

UNIVERSIDAD CARLOS III DE MADRID

Máster en Bibliotecas y Servicios de Información Digital

Una aproximación al data curation y el rol del bibliotecario en su implantación

El caso de la Universitat Oberta de Catalunya

Mireia Pérez Cervera

Octubre 2012

Tutores: Gema Bueno de la Fuente y Jesús Robledano Arillo

Sumario

Sumario	1
Índice de figuras	4
Índice de tablas	4
Resumen ejecutivo	5
1. Introducción	6
1.1. Justificación	6
1.2. Objetivos	7
1.3. Metodología	8
1.3.1. Fase bibliográfica: construcción del marco teórico	8
1.3.2. Fase bibliográfica: análisis de recomendaciones y ejemplos de buenas prácticas ..	9
1.3.4 Fase proyectiva	9
1.4 Alcances y limitaciones	10
1.5 Fuentes consultadas	11
2. Data Curation: estado de la cuestión	14
2.1. La evolución de la ciencia. La evolución de los datos	14
2.2. El concepto de <i>Data Curation</i>	16
2.3. Principales organismos e instituciones de <i>data curation</i>	22
2.4. Un nuevo reto para las bibliotecas	28
2.5. Bibliotecas en acción: algunos ejemplos	30
2.6. El papel del bibliotecario	31
3. Herramientas y recursos para la preservación de datos de investigación	38
4. Informe de la situación del <i>data curation</i> en la Universitat Oberta de Catalunya	45
4.1. Introducción: la Universitat Oberta de Catalunya	46
4.1.1. La institución	47
4.1.2. La Biblioteca	49
4.1.3. La actividad investigadora	51
4.2. <i>Data curation</i> en la UOC	54
4.2.1. Comentarios previos	54
4.2.3. Marco de regulación legislativo y el papel de las organizaciones	55
4.2.4. Gestión de los datos de investigación en la UOC	57
4.2.5. Tecnología UOC y los datos de investigación	58
4.2.6. Agentes implicados	59
4.2.7. Experiencias externas de <i>data curation</i>	60

4.2.8. Análisis DAFO.....	61
5. Propuestas de actuación.....	63
6. Conclusiones.....	69
Bibliografía.....	70
Anexo: Listado de instituciones, organismos y páginas web consultadas	81

Índice de figuras

Figura 1. DCC Curation Lifecycle Model	19
Figura 2. Ciclo de los datos por el UK Data Archive [http://www.data-archive.ac.uk/create-manage/life-cycle].....	211
Figura 3. Políticas de datos de los principales organismos de financiación. Fuente: DCC ...	27
Figura 4. La pirámide de Lewis.....	33
Figura 5. Perfil híbrido del profesional de la información (Corall 2008).....	36
Figura 6. Core Skills for Data Management, de Donnelly (2008).....	37
Figura 7. Estructura del informe para la UOC	46
Figura 8. Campus Virtual UOC. Página de inicio.....	48
Figura 9. Recursos de la Biblioteca Virtual UOC	620
Figura 10. Mapa de conocimiento de la UOC.....	623
Figura 11. DAFO de los datos de investigación en la UOC.....	62
Figura 12 Factores críticos del Plan Director.....	64

Índice de tablas

Tabla 1. Recursos y herramientas para la gestión de datos de investigación	38
--	----

Resumen ejecutivo

Este trabajo trata de la gestión, recuperación y preservación de los datos de investigación en el seno de una institución, y del papel que desarrolla la biblioteca en este proceso.. Se divide en cinco partes, las tres primeras analizan el significado del término de *data curation*, y las definiciones propuestas por distintos autores con la finalidad de comprender la aplicación de cada una de las fases . Se realiza una prospección general sobre las principales entidades internacionales que están trabajando en este ámbito, y que con sus actividades y recursos realizan una destacada labor de apoyo, liderazgo y formación para el bibliotecario y el investigador.

Una vez adquiridos los conocimientos básicos sobre la aplicación de sistemas de *data curation*, se procede a afrontar la situación de un caso particular. Se realiza un informe diagnóstico sobre la Universitat Oberta de Catalunya, para conseguir, mediante el análisis de la gestión de los datos y el funcionamiento de esta institución, elaborar la propuesta de un conjunto de medidas y recomendaciones que mejoren la situación de los datos de investigación.

Como resultado, se obtiene una fotografía general de la situación del *data curation*, así como un documento de gestión interna para la UOC que, con un claro apoyo institucional podrá ser aplicado a corto o medio plazo.

1. Introducción

1.1. Justificación

Nos encontramos en la era de la sociedad de la información en la que ésta forma parte de aspectos muy diversos de la vida económica, social y cultural. Los avances tecnológicos junto con el crecimiento exponencial del acceso a la información que se ha producido durante las últimas décadas están produciendo enormes cambios en también en el seno de la actividad científica. En este sentido, la actividad de investigación ha evolucionado mediante la combinación de estudios de campos de conocimiento, hasta la forma de conducta de los investigadores, quienes tienen ahora otra oportunidad para ampliar y mejorar sus investigaciones.

Desde la perspectiva de los profesionales de la información que trabajan ofreciendo apoyo a la investigación se plantean nuevos retos. Este trabajo presta atención a los cambios que están afectando a los distintos procesos de gestión y preservación de los datos de investigación y a la necesidad de adaptar a esta nueva situación los roles de los profesionales e la información. Los datos de investigación que se van generando durante todo el proceso de la actividad científica, están experimentando cambios profundos en relación con el número de datos creados, procesados y diseminados por los investigadores como consecuencia de su actividad profesional (National Research Council , 2009).

El presente trabajo surge de la advertencia de una necesidad que profesionalmente y desde la UOC se comparte con muchas instituciones, universidades y profesionales del mundo de la información y documentación en relación con la gestión de la preservación de datos que se generan durante todo el proceso de investigación, lo que en la terminología inglesa se conoce como *data curation*.

El trabajo propone la realización de un estudio de la situación de una institución académica en relación con la gestión de datos de investigación (*research data curation*), concretamente en el caso específico de la Universitat Oberta de Catalunya (UOC). Para ello se reflexiona sobre el actual contexto de la preservación

de datos, se analiza las posibilidades de actuación en la UOC, y se definen algunas propuestas para mejorar la preservación de datos de investigación.

Asimismo, la elección de este tema se justifica por el interés de la autora por construir una base teórica y práctica que pueda aplicarse a su contexto laboral, la Universitat Oberta de Catalunya, que en la actualidad carece de cualquier política o gestión de datos de investigación.

1.2. Objetivos

El objetivo principal de este trabajo es la creación de un marco teórico y práctico general sobre la preservación de los datos de investigación que ayude a los profesionales de la información y a los investigadores de la Universitat Oberta de Catalunya a entender el concepto de *data curation* y a contextualizarlo en su entorno actual.

Se propone:

- ❖ Realizar una aproximación al tema del *data curation*, de forma que se asimile de manera comprensible el significado de la preservación de datos de investigación.
- ❖ Obtener una visión sintetizada del contexto general del *data curation*, en relación con los organismos que lo apoyan, y las herramientas disponibles.
- ❖ Elaborar un informe que establezca como poner en marcha un proyecto de mejora de los datos de investigación de una institución académica, ejemplificándolo en la UOC.
- ❖ Realizar un conjunto de propuestas que puedan ser aplicadas por la biblioteca de la UOC para mejorar la gestión y la preservación de los datos de investigación sin la necesaria ejecución de un plan de *data curation*.

Se plantea la creación de un informe que mediante la realización de fases de análisis previas permita llevar a cabo el desarrollo de algún servicio o mejora de la gestión de datos de investigación en la Universitat Oberta de Catalunya. El objetivo

último es la consecución de un documento teórico y práctico que sirva a la los bibliotecarios del grupo de apoyo a la investigación de la UOC, plantear un plan director para el desarrollo de actividades de *data curation* en su institución.

1.3. Metodología

Para la consecución de los objetivos establecidos para el presente trabajo, se realiza una primera descripción del actual contexto de crecimiento de los datos de investigación y del problema de su preservación por parte de las instituciones y personas implicadas. Posteriormente, se proponen unas pautas básicas y se lleva a cabo un análisis previo que permita establecer las bases de la situación de la UOC en relación con la preservación de datos de investigación, y facilitar, en un futuro, un plan de actuación de *data curation*.

1.3.1. Fase bibliográfica: construcción del marco teórico

Para iniciar este estudio se ha seguido, como primera fase metodológica, la lectura y análisis de diversos títulos de bibliografía básica sobre gestión de datos de investigación, en base a recomendaciones de organismos internacionales o de personalidades reconocidas del *data curation*, así como una selección de artículos de revistas de renombre de ámbito internacional con la finalidad de asimilar la definición de la gestión de datos como los distintitos procesos que su realización supone.

Para la selección de esta bibliografía se tuvo especialmente en cuenta el listado bibliográfico que recientemente publicó Charles W. Bailey. El listado completo sobre *Research Data Curation Bibliography* (Bailey, 2012) Se trata de una selección de bibliografía muy reciente y actualizada dada la relativa novedad del tema. El artículo más antiguo de la selección de Bailey, data del 2004, se conoce que el tema del *data curation* está de plena actualidad, y en estos momentos están apareciendo continuamente nuevos trabajos, por lo que nos enfrentamos a un campo de investigación en pleno proceso de emergencia.

1.3.2. Fase bibliográfica: análisis de recomendaciones y ejemplos de buenas prácticas

La segunda fase de documentación bibliográfica ha consistido, por un lado, en la consulta de ejemplos de casos prácticos de implementación de sistemas de *data curation* en instituciones de investigación. Para ello se ha realizado una revisión de la documentación sobre organismos, instituciones, congresos y otras entidades relacionadas con la investigación y desarrollo del *data curation*. Por otro lado, y tras la asimilación de los conceptos teóricos básicos del *data curation*, se analiza el papel de la biblioteca académica, el rol de los bibliotecarios respecto el *data curation*, y los recursos y herramientas disponibles en la actualidad y un análisis de documentación institucional existente, (guías, recomendaciones, ejemplos de buenas prácticas). En este sentido, se analizan los casos reconocidos internacionalmente (*Massachusetts Institute of Technology, Data Curation Center* o *DataToolkit*).

1.3.4 Fase proyectiva

En esta fase se realiza una propuesta de los requerimientos básicos que cualquier institución académica que se plantee asumir la gestión de datos y emprender acciones de *data curation* debería examinar y analizar para determinar la viabilidad y contexto del desarrollo de un plan de gestión de datos.

Para desarrollar la parte más práctica de esta fase, se toma como referencia la biblioteca de la Universitat Oberta de Catalunya. Los datos para este apartado han sido recopilados mediante la consulta de su página web así como blogs, recursos internos y webs de distintos grupos de investigación. También se ha tenido acceso directo a varios documentos de gestión interna, la mayoría de ellos realizados por el equipo de la biblioteca virtual, como informes de la actividad científica, informes de proyectos o entrevistas. Se ha consultado la documentación alojada en el campus virtual de la universidad, especialmente aquellos apartados relativos a la

actividad de investigación de la institución y la intranet del equipo de gestión. Para la redacción de esta parte del trabajo se han seguido las rutinas habituales de trabajo interno en la UOC. Además se ha podido acceder a documentación e informes internos que incluían valiosos comentarios y resultados de la comunicación directa con algunos de los investigadores de la institución mediante entrevistas definidas por un equipo interdisciplinar de la biblioteca.

1.4 Alcances y limitaciones

El campo de los datos de investigación puede ser tratado desde distintos puntos de vista, como los conceptos de *open access*, el uso de aplicaciones de acceso libre, o del *data sharing*, que implica compartir los datos de investigación con otras personas (Torres-Salinas 2010). Hay varias cuestiones relacionadas con el tema que nos ocupa, todas ellas de gran interés, pero los límites de este trabajo no nos van a permitir abarcarlas todas.

Por ello, cuando hablemos de *data curation* nos referiremos al tratamiento, gestión y preservación de datos estrictamente relacionados con los resultados de procesos de investigación. Para el desarrollo del proyecto centrado en los datos de la Universitat Oberta de Catalunya, solamente se van a tener en consideración los datos de la investigación de esta universidad, no se contemplan datos administrativos o creados por la institución.

Así mismo se pretende insistir en la distinción de los conceptos de *data curation* y de *open data*, el movimiento a favor del libre acceso a los datos sin restricciones legales de cualquier tipo, para su uso y redistribución (Murray-Rust, 2008). Aunque se trate de campos muy cercanos y a menudo relacionados, este trabajo quiere centrarse en la gestión de los datos de investigación, sin participar de los debates que puedan plantearse sobre su apertura o no. Los conceptos de *open data* y *data sharing*, acostumbran a ir de la mano cuando se habla de gestión de datos y repositorios. En cualquier caso, aunque esta cuestión pueda aparecer en el planteamiento de cualquier proyecto de *data curation*, en este trabajo no va a ser el

principal objeto de debate, sino la gestión de los datos como profesionales de la información.

1.5 Fuentes consultadas

Las fuentes utilizadas que permiten construirse una idea sobre el estado de la cuestión, informar de las diferentes corrientes actuales y conocer los procesos a seguir para una correcta gestión y preservación de datos. En este sentido tanto la bibliografía de Bailey como la facilitada por *Digital Curation Bibliography* han resultado la base para recopilar información sobre las distintas etapas y actividades relacionados con los procesos de *data curation*. Se trata de una selección de bibliografía muy reciente y actualizada dada la relativa novedad del tema. Las páginas oficiales de organismos especializados o empresas como GrandIR¹, *Digital Curation Centre* (DCC), *Massachusetts Institute of Technology* (MIT) o *National Academy of Science* (NAS) permiten formarse una idea de las pautas y bases de desarrollo del *data curation* en la actualidad. De esta manera se analiza el contexto español (GrandIR), el europeo (DCC) o la visión norteamericana (MIT y NAS). Conocer estos diferentes estadios de la misma cuestión favorece el establecimiento de una posible graduación en la evolución de nuestro proyecto. También han resultado interesantes las lecturas de las actas y ponencias de un congreso organizado por la Biblioteca Nacional de Holanda en el 2009 (Angevaare, 2009) y el Open Repositories organizado por la *University of Edinburgh* en 2012 .

Han sido dos los manuales que han permitido la conceptualización práctica y profesional *del data curation*. La obra de Ross Harvey (2010) hace hincapié en pautas y recomendaciones de carácter práctico para implantar un proyecto de *Digital curation*; mientras que la obra coordinada por Graham Pryor (2012) director asociado de DCC, tiene un carácter más teórico y reflexivo sobre elementos críticos de la gestión de datos y sobre estrategias y planes de desarrollo de datos.

¹ Las URL de los organismos e instituciones que se citan en este trabajo se incluyen al final del mismo, en un glosario de sitios web e instituciones.

Junto a los manuales, algunos de los trabajos que han resultado de especial interés han sido Barateiro y otros (2008), Joyce Ogburn (2010), o el informe de la *National Academy of Sciences Committee* (National Research Council 2009) que tratan de los datos y formulan algunos principios básicos para asegurar su integridad, acceso y custodia. Por su parte, los artículos de Gutmann y otros (2009), Heidorn (2011), Joint (2007), Angevaere (2009) y Ramírez (2011) contribuyen a un mejor conocimiento de los nuevos retos de los bibliotecarios y bibliotecas académicas en relación con la gestión de datos de investigación, mientras que el documento de Tenopir (2011) enfoca el tema desde el punto de vista de los investigadores. Finalmente los trabajos de Macdonald (2010), Wilson (2011) y Beagrie (2008) aportan ideas sobre el análisis de la situación de la gestión de los datos de investigación en la UOC.

Estos representan los principales documentos que han ayudado a la elaboración de este trabajo, pero no los únicos. Se podrá consultar la relación bibliográfica completa al final del documento.

1.6 Estructura del trabajo

En la primer parte se trata el concepto de *data curation*, sus actividades básicas y el ciclo de vida de los datos a partir de la revisión de la literatura más reciente y relevante. En la segunda parte se analizan casos de buenas prácticas e iniciativas de organismos e instituciones implicadas en la preservación de datos de investigación. En este sentido, se reflexiona sobre el reto que el data curation representa para bibliotecas y profesionales de la información. En el tercer apartado se realiza un análisis de las principales herramientas y recursos que existen en la actualidad para afrontar dicha actividad, se enumeran y se comparan las prestaciones que ofrecen. En el cuarto apartado, el trabajo se centra en el caso real de la Universitat Oberta de Catalunya (UOC). Se analiza su situación en relación con los datos de investigación, mediante el análisis de su contexto interno y un DAFO, en el quinto y último apartado se proponen pasos a seguir para que se

realice una mejora de esta actividad y se proponen acciones concretas para que puedan ser implementadas mediante la elaboración de un Plan Director en esta universidad.

2. Data Curation: estado de la cuestión

*“A fundamental characteristic of our age
is the rising tide of data – global, diverse,
valuable and complex. In the realm of
science, this is both an opportunity and a
challenge.”*

(High level Expert Group on Scientific Data, 2010)

2.1. La evolución de la ciencia. La evolución de los datos

La actividad investigadora ha estado sujeta a grandes cambios durante la historia de la ciencia. Especialmente cabe destacar los sufridos a partir del siglo XXI a raíz de la llegada de las nuevas tecnologías. Las metodologías tradicionales de la investigación, basadas en la experimentación y la teoría de la hipótesis, han sido substituidas por nuevos protagonistas estrechamente vinculados a la era digital.

Por un lado, el desarrollo de la e-ciencia (*e-science* o *enhanced science*), que según Martínez Uribe y Macdonald “cuando la investigación multidisciplinar y en colaboración tiene lugar en diversas localizaciones, produciendo y utilizando grandes cantidades de datos” (Martínez-Uribe, y Macdonald, 2008). Por otro lado, existe la influencia de una nueva e-infraestructura, el nuevo entorno de servicios e información tecnológica sobre el que se apoya el trabajo y la investigación de nuestra sociedad de la información. La conjunción de e-ciencia y de e-infraestructuras permite la aparición de nuevos métodos y herramientas de investigación contribuyendo a una mayor interdisciplinaridad, a una apertura de las fronteras geográficas y a la llegada de nuevas formas de almacenar y gestionar la información.

Durante las últimas décadas los datos en formato electrónico han sido el principal medio de producción científico y, aunque esta situación se ha hecho patente en todos los campos de investigación, resulta especialmente acusada en el ámbito de las ciencias donde el ratio de crecimiento ha llegado a una cifra cercana al 30%.

Esta situación ha propiciado la aparición del término *digital deluge* (el diluvio digital) que no solamente define los datos directamente generados por nuevos proyectos nacionales e internacionales de investigación sino que también define el incremento que estos datos reciben en su posterior reutilización en todo el mundo (Pryor, 2012).

Como apunta Borgman (2007), los datos están empezando a ser valorados como un producto académico final, es un momento de puesta en valor de los datos científicos. En el informe del National Research Council (2009) se encuentran premisas sobre la gestión de datos creados en este nuevo contexto tecnológico que enfatizan la importancia de la protección de la integridad, accesibilidad y custodia de los datos de investigación. Para ello se observa la necesidad de desarrollar políticas, estándares e infraestructuras que requieren el esfuerzo por parte de todos sus participantes: investigadores, instituciones, organismos y bibliotecas.

Los beneficios que pueden impulsar la decisión de emprender programas de *data curation* en distintos ámbitos son múltiples y afectan a diferentes actores: a los ciudadanos, porque obtienen los resultados de sus contribuciones de impuestos, y porque reciben la transferencia de la ciencia a la sociedad; a las agencias de financiación de investigaciones porque se facilita la toma de decisiones basadas en las evidencias, se evita la duplicidad del trabajo y se recuperan inversiones; los investigadores porque un sistema de *data curation* facilita la consulta de todos sus datos incrementando su productividad; y a las grandes industrias y empresas porque utilizan esta información para aumentar su productividad, crear nuevo conocimiento, mercados, oportunidades de trabajo e incrementar las posibilidades de movilidad e intercambio de conocimiento (High level Expert Group on Scientific Data, 2010).

2.2. El concepto de *Data Curation*

La aparición del concepto de *data curation* responde a la necesidad de encontrar una solución a la correcta gestión del ingente volumen de datos provenientes de la investigación científica. Pese a que durante algún tiempo existiese la equívoca tendencia de utilizar el concepto de “data” como sinónimo de “digital data”, a día de hoy cuando se habla de *data curation* se hace en el sentido más amplio del término.

El término anglosajón *data curation* se utiliza para referirse a las acciones necesarias para el mantenimiento y utilización de los datos digitales y los resultados de investigación durante todo su ciclo de vida y desde su generación. El término lleva implícitos procesos como el archivo o la preservación de datos, así como todos aquellos procesos necesarios para la creación y mantenimiento de los datos, y la capacidad de aportar valor para generar nuevas fuentes de información y conocimiento.

En cambio en el ámbito español no existe una única terminología del concepto de *data curation*, y se utilizan diferentes denominaciones, como curación de datos o preservación de datos de investigación (Keefer, 2011). A lo largo de este trabajo, se utilizará la denominación anglosajona *data curation* para referirse a la preservación de los datos de investigación y para hablar de las mismas actividades que prevé el término original.

El Digital Curation Center (DCC), organismo británico al que nos referiremos más adelante, define *data curation* como aquella actividad que se centra en la conservación, preservación y la adjudicación de valor añadido a los datos de investigación durante todas y cada una de las fases de su ciclo de vida:

“Digital curation is maintaining and adding value to a trusted body of digital research data for current and future use; it encompasses the active management of data throughout the research lifecycle (Data Curation Center, 2012).”

El DDC gestiona y promueve el uso de los datos desde el momento de su creación para asegurar su permanencia y disponibilidad para ser localizados y reutilizados, minimizando así los riesgos de desaparición y obsolescencia.

Ésta es solamente una de las muchas definiciones de *data curation* que existen, la que realiza Ross Harvey amplia los conceptos apuntados por el DCC:

“Digital curation is concerned with actively managing data for as long as it continues to be scholarly, scientific, research, administrative, and/or personal interest, with the aims of supporting reproducibility, reuse of and adding value to that data, managing it from its point of creation until is determined not to be useful, and ensuring its long-term accessibility, preservation, authenticity, and integrity (Harvey, 2010).”

Según este autor las principales características de la actividad de *data curation* son el conjunto de procesos aplicados a objetos digitales durante todo el transcurso de su vida, desde su creación hasta su almacenamiento final.

2.3. Los datos. El ciclo de vida.

Cada uno de los datos en los que se generan durante la investigación adquiere una entidad propia y son reutilizados en futuras investigaciones. Se consideran datos de investigación en formato digital aquellos que son creados durante el proceso de investigación y que se almacenan en cualquier formato digital.

El *National Science Board* (NSB) de los EEUU define los datos de investigación de la siguiente manera:

“Any information that can be stored in digital form, including text, numbers, images, video or mobiles, audio, software, algorithms, equations, animations, models, simulations, etc. Such data may be generated by various means including observation, computation or experiment”. (National Science Board, 2005)

La diversidad de tipologías con la que nos enfrentamos es enorme, lo que agrava la dificultad de su gestión. El término “datos de investigación” se refiere a una gran variedad de datos, que a su vez se disponen de diferentes formatos:

- Datos numéricos: resultados de medidas de instrumentos, resultados de encuestas.
- Documentación: publicaciones, notas de investigación de campo, descripciones, informes.
- Imágenes digitales: fotografías y diagramas
- Audio digital: archivos sonoros, entrevistas.
- Vídeo digital: emisiones, entrevistas.

Cada uno de ellos también pueden ser tratados, procesados o bien utilizados en crudo.

Los riesgos de la falta de gestión de datos son la creación de datos de forma indiscriminada y sin normalización ni control institucional, lo que conlleva un altísimo riesgo de pérdida de información a largo plazo, de escasa reutilización e interoperabilidad. Estos inconvenientes están directamente relacionados con el aumento del coste de la obtención de datos y con una escasa transparencia de los resultados científicos.

Para una correcta gestión de los datos de investigación, es decir, para asegurar su accesibilidad y disponibilidad a lo largo del tiempo, resulta necesaria una planificación, por lo que distintos organismos han propuesto una gestión documental de los datos basándose en su ciclo de vida. Esta forma de administración de datos durante su ciclo de vida permite asegurar que se desarrollen políticas y procesos, se definan roles y responsabilidades y se cree un marco tecnológico adecuado para su almacenamiento (Pryor, 2012).

El modelo de la DCC tiene como principal objetivo la preservación de los datos y ofrece una visión ideal de las actividades que deberían llevarse a cabo, así como las políticas, procedimientos, roles y responsabilidades, estándares y tecnologías a desarrollarse.

En esta propuesta, el ciclo de los datos comienza con la descripción y la representación de la información. Este primer punto puede resultar problemático dada la variedad de formatos que pueden adquirir los datos de investigación. En esta primera etapa será necesario colaborar muy de cerca con los creadores de los datos para comprenderlos bien. La vigilancia de la comunidad y la participación dependerán de cada grupo de datos y sus usuarios. Las comunidades a las que se refiere serán las encargadas de desarrollar estándares y prácticas. Conservar y preservar son las actividades a llevar cabo a través del plan de preservación anteriormente definido.

La figura 1 muestra las fases que a nivel conceptual deben seguirse para llevar a cabo un proceso de preservación de datos:

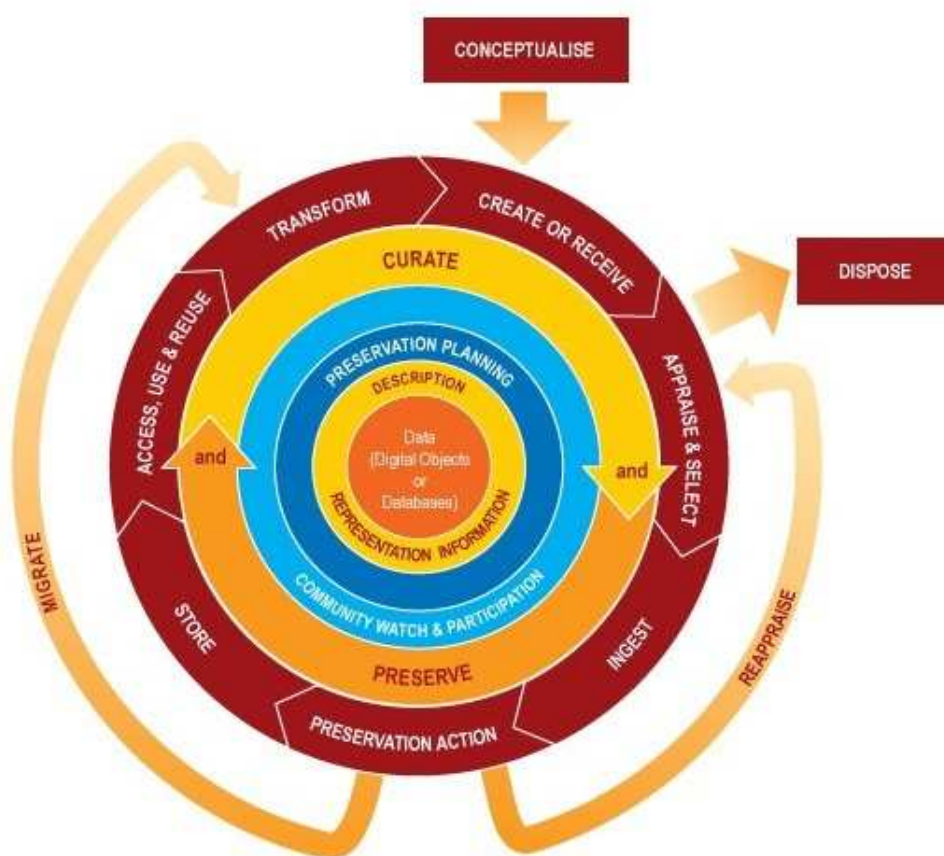


Figura 1. DCC Curation Lifecycle Model

Acciones ligadas al proceso de almacenamiento de datos partiendo del apartado de conceptualización:

- *Crear o recibir (create or receive)*: Es en el inicio del ciclo de los documentos donde empieza la actividad del *data curation*. Se trata del proceso de generación de datos por parte del investigador, o bien del proceso de recepción de datos procedentes de otras colecciones de fuentes de datos. Es en esta etapa inicial de la vida de los datos en la que el rol del bibliotecario resulta vital para aportar apoyo a los investigadores y asegurar que los datos son recogidos en el formato apropiado y que son descritos con los metadatos correctos.
- *Evaluar y seleccionar (appraise and select)*: en este apartado, en base a las políticas y recomendaciones previamente establecidas, se evalúan y seleccionan los datos que serán preservados a largo plazo.
- *Desechar (dispose)*: los datos que no son previamente seleccionados en concordancia con la política de selección se deben desechar.
- *Traspaso (ingest)*: proceso de transferencia de los datos de investigación a un repositorio, archivo, centro de datos...
- *Acción de preservación (preservation action)*: esta etapa comprende todas aquellas acciones que se deben tomar para asegurar la correcta preservación de los datos, preservar su autenticidad, integridad y reusabilidad. Los datos digitales pueden estropearse fácilmente, para evitarlo y asegurar su preservación se deben revisar errores, formatos, realizar migraciones, copias de seguridad, etc.
- *Almacenar (store)*: proceso de almacenamiento de la información.
- *Acceso, uso y reutilización (access, use and reuse)*: la finalidad de esta etapa es la de facilitar la localización de los datos, de forma que se deberán añadir metadatos o indexar datos.
- *Transformación (transformation)*: la etapa final del ciclo de vida de los datos propuesto por la DCC incluye la necesidad de la transformación de los datos a través del tiempo en distintos formatos para evitar la obsolescencia de los

software. Mediante esta acción los datos vuelven a encontrarse al inicio de su ciclo de vida.

El ciclo de vida de los datos de investigación solamente se comprende en el seno de un marco institucional dotado de profesionales con conocimientos y habilidades para verificar grandes sumas de datos, transformar los formatos de los documentos y crear scripts que permitan la transferencia de grandes archivos de datos a terceros (National Research Council, 2009).

Otro de los organismos que ofrece una propuesta de ciclo de vida de los documentos es el UK Data Archive, definiendo las seis etapas básicas para la preservación de los datos de investigación y dirigiéndose especialmente al usuario investigador:

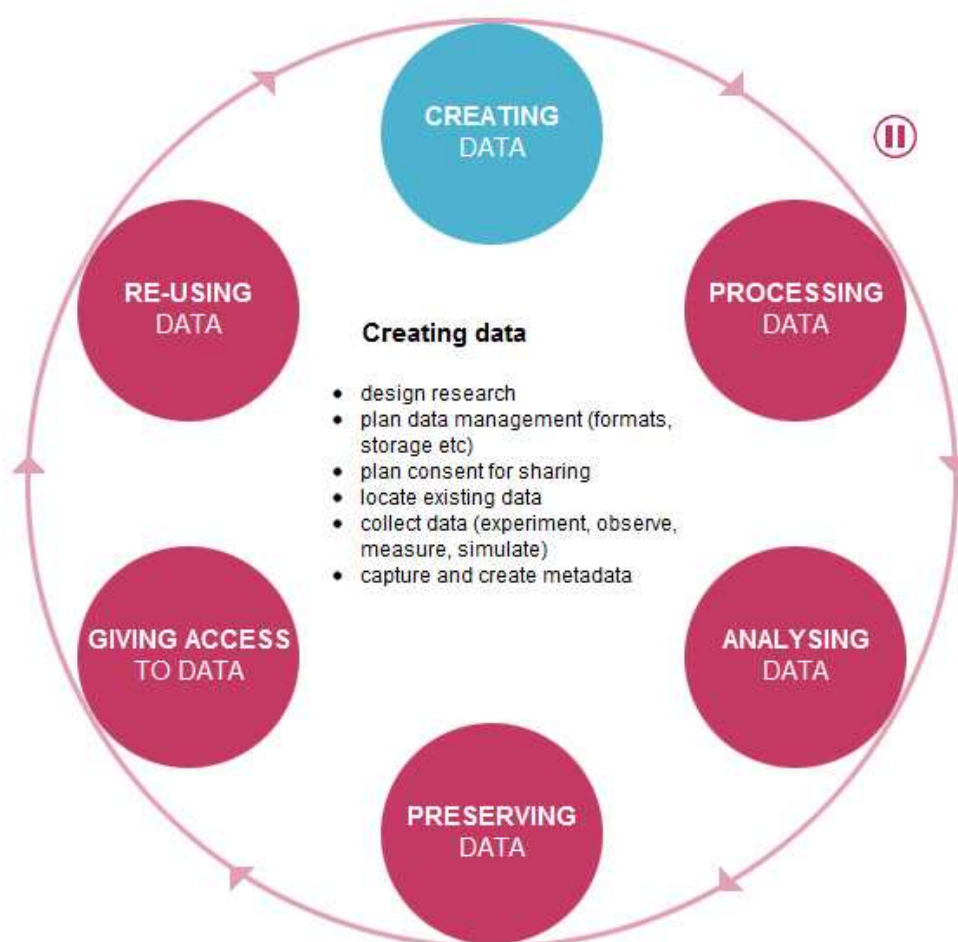


Figura 2. Ciclo de los datos por el UK Data Archive [<http://www.data-archive.ac.uk/create-manage/life-cycle>]

Las etapas definidas en este modelo comprenden las acciones siguientes:

- *Creación de datos (creating data)*: diseño de la investigación, creación de planes de gestión de datos, políticas de data sharing, localización de los datos, recolección y creación de metadatos.
- *Procesamiento de datos (processing data)*: introducción, digitalización, transcripción, traducción, validación, limpieza, descripción, almacenamiento y protección de datos (en los casos en que deba asegurarse el anonimato).
- *Análisis de los datos (analysing data)*: interpretación y preparación de los datos para su preservación.
- *Preservación de los datos (Preserving data)*: migración de los datos al formato recomendado, creación de backups, creación de metadatos y documentación, archivo de los datos.
- *Dar acceso a los datos (giving Access to data)*: distribuir, compartir y controlar el acceso a los datos, como establecer licencias y promocionar los datos.
- *Reutilizar los datos (reusing data)*: realizar un seguimiento de la investigación, retomar nuevos resultados, enseñar y aprender de ellos.

2.4. Principales organismos e instituciones de *data curation*

La creación de organismos, instituciones y centros de investigación que han centrado sus esfuerzos en desarrollar actividades, guías, y en definitiva, en el estudio sobre la preservación de datos de investigación, es bastante reciente, sobretodo desde principios del s.XXI. Los más relevantes se encuentran ubicados en países anglosajones como Reino Unido, Estados Unidos y Australia.

De entre los principales organismos e instituciones ligadas a la actividad de *data curation*, distinguimos dos tipos. El primero lo conforman los centros de datos y los repositorios, quienes resultan de especial importancia, puesto que no solamente proveen de información y orientan en la gestión de los datos, sino que soportan los

costes necesarios para el mantenimiento de infraestructuras necesarias para el almacenamiento, preservación y acceso a los datos. El rol de estos organismos resulta de mayor importancia cuando se trata de centros nacionales, puesto que reciben el apoyo y financiamiento gubernamental y se convierten en centros de referencia de un país.

El segundo tipo serían las instituciones de financiación de la investigación. Estos organismos son los que disponen de la potestad para pedir a los investigadores un especial trato en la preservación de sus datos como requerimiento para su financiación, y acostumbran a desarrollar documentos de buenas prácticas muy útiles para los creadores de datos. Los centros de datos han tenido un crecimiento especial en el contexto del Reino Unido, donde la mayoría de ellos son financiados, a su vez, por alguno de los concilios de investigación o por el *Joint Information Systems Committee* (JISC) (Pryor, 2012). Son muchos los centros de datos que existen en la actualidad, y abarcan una gran diversidad de campos de investigación: arqueología, ciencias atmosféricas, química, bioinformática, ciencias de la tierra... En este apartado solamente se resaltan los más significativos.

2.4.1 Centros de datos y repositorios de datos

A continuación se describen algunos de los centros y repositorios de datos que a nivel internacional pueden ofrecer una visión general de lo que se está realizando a este nivel. No se trata de una selección exhaustiva, sino representativa.

- **Digital Curation Center (DCC)**

El DCC es uno de los organismos más relevantes en la preservación de datos en el Reino Unido y que se ha erigido como referente a nivel nacional e internacional. Fue creado en el año 2004 como centro de preservación digital con los objetivos de crear, capacitar y dotar de herramientas para la gestión de datos de investigación a la comunidad científica de las instituciones de educación superior en el Reino Unido. El DCC realiza variadas actividades, desde la creación de recursos y

programas de formación hasta el apoyo en la creación de políticas de desarrollo planes de gestión de datos. Resultan especialmente interesantes las herramientas y materiales que han generado para crear planes de gestión de datos de investigación y que pueden descargarse directamente desde su página web.

UK Data Archive (UKDA)

Uno de los organismos más importantes hasta la fecha es el *UK Data Archive* (UKDA), una organización nacional británica que funciona como centro de datos. Este centro de datos de ciencias sociales y humanidades funciona desde 1963, y fue una de las principales instituciones que reconoció la importancia de realizar una recogida de datos a nivel nacional. El UKDA ofrece información a los investigadores y facilita información sobre los temas de confidencialidad, copyright, metadatos, formatos y software.

Arts and humanities Data Service (AHDS)

El *Arts and humanities Data Service* (AHDS) fue un organismo nacional fundado por el *JISC* y el *Art and Humanities Research Council* el 1995 con la finalidad de recolectar, preservar y promover los recursos electrónicos creados como resultado de las actividades de investigación y educación de los campos de artes y humanidades. Esta institución realizaba un papel relevante sobre la preservación de datos en campos de investigación de humanidades, llevando también a cabo diferentes formaciones, orientación, guías y manuales temáticos. Fue en 2008 cuando dejó de recibir subvenciones estatales complicándose así su supervivencia y viéndose obligado a cesar sus actividades. Durante 12 años desarrolló en el ámbito del Reino Unido un papel de difusión y formación muy importante.

Australian National Data Service (ANDS)

El *Australian National Data Service* lleva más de diez años elaborando programas para la transformación de las colecciones de datos de investigación australianos en una red de repositorios de investigación, y con una actividad paralela en la formación de expertos. En los últimos tiempos el ANDS está centrando sus esfuerzos en cuatro puntos estratégicos para transformar los datos en colecciones estructuradas: de lo no gestionado a la gestión; de lo desconectado a lo conectado; de lo invisible a lo visible; del uso único a lo reutilizable. El objetivo del proyecto que en la actualidad está llevando a cabo va a ser la consecución de una herramienta para los investigadores australianos que les faciliten la publicación, acceso y uso de los datos de investigación del país (Pryor, 2012)

Archaeology Data Service (ADS)

Originariamente este organismo formaba parte del AHDS, fue tras su desaparición que se erigió como institución independiente. De entre sus actividades destaca la elaboración de una guía para depositar datos arqueológicos, la promoción de buenas prácticas en el uso de datos digitales de arqueología, el asesoramiento técnico a la comunidad de investigación y el apoyo a la implementación de tecnologías digitales.

Aparte de los centros y repositorios de datos, existen otras instituciones relacionadas con la preservación de datos de investigación que merecen una cita, por su peculiaridad y por lo que ofrece en relación al *data curation*:

Dryad

Dryad es un depósito internacional de datos científicos, principalmente procedentes de artículos de investigación del campo de la biociencia. Se trata de un proyecto está siendo desarrollado por el *National Evolutionary Synthesis Center* y el centro de investigación de metadatos de la Universidad de Carolina del Norte, y

gestionado por un consorcio de revistas que promueven colaborativamente el archivo de datos y garantizan la sostenibilidad del repositorio.

3TU.Datacentrum

Esta institución funciona mediante la colaboración de tres universidades técnicas holandesas (TU Delft Library, TU Eindhoven Information Expertise Centre y University of Twente Library & Archive) y facilita un centro de investigación para el archivo de datos. Este organismo ofrece servicios que incluyen programas de preservación a largo tiempo, acceso permanente a los datos y a su reutilización y apoyo en la formación y gestión de datos.

Network of expertise in Long-Term Storage (NESTOR)

Esta institución tiene como objetivos asegurar la preservación de los recursos digitales en Alemania y trabajar con otras instituciones internacionales para asegurar su memoria digital global. Forma una red de competencias para la preservación y ha definido estándares para repositorios digitales.

2.4.2. Organismos de financiación

El *Digital Curation Center* elaboró un informe (Jones 2009) muy completo sobre la situación de los centros de datos a nivel internacional, que resumió en la siguiente tabla, en la que se puede ver el tipo de actividades que cada uno de ellos lleva a cabo en relación con las políticas de datos y el apoyo que ofrecen para su conservación:

● Full Coverage
 ◐ Partial Coverage
 ○ No Coverage

Research Funders	Policy Coverage		Policy Stipulations					Support Provided			
	Published outputs	Data	Time limits	Data plan	Access/sharing	Long-term curation	Monitoring	Guidance	Repository	Data centre	Costs
AHRC	●	●	●	●	●	◐	○	●	○	◐	○
BBSRC	●	●	●	●	●	●	●	●	●	◐	●
CRUK	●	●	●	●	●	●	●	◐	●	○	○
EPSRC	●	●	●	◐	●	●	●	◐	○	○	●
ESRC	●	●	●	●	●	●	●	●	●	●	◐
MRC	●	●	●	●	●	●	○	◐	●	○	◐
NERC	●	●	●	●	●	●	●	●	●	●	◐
STFC	●	●	●	●	●	●	●	◐	●	◐	○
Wellcome Trust	●	●	●	●	●	●	●	●	●	◐	●

Figura 3. Políticas de datos de los principales organismos de financiación. Fuente: DCC

Se han seleccionado algunos centros de datos representativos, algunos de ellos, como el Arts and Humanities Research Council (AHRC) y el Biotechnology and Biological Sciences Research Council en base al análisis del DCC y el National Science Foundation (NSF) que abarca la experiencia americana.

El Arts and Humanities Research Council apoya a la investigación de las humanidades que “promueven el entendimiento de la cultura humana y la creatividad”. Ha desarrollado una declaración sobre acceso abierto a la investigación en línea con el Research Council del Reino Unido (RCUK) en el que se especifica la necesidad de los investigadores de preparar un plan de gestión de datos, en el que se abordan los siguientes temas: aspectos técnicos; métodos de desarrollo de datos; infraestructura; preservación de datos y sostenibilidad; acceso; derechos de autor y derechos de propiedad intelectual. Destaca el hecho que el AHRC requiere a sus investigadores la obligatoriedad de conservar los conjuntos de datos más importantes en un repositorio accesible durante al menos tres años. Solamente en el caso de los datos arqueológicos se contempla una preservación a largo plazo, la cual delegan a otra institución, la Archeology Data Center. Por otro lado el *Biotechnology and Biological Sciences Research Council* es

un organismo que no prescribe requisitos obligatorios para la preservación de los datos de investigación. Aún así, aconseja a sus investigadores que pongan los datos accesibles de una “manera oportuna y responsable” para su reutilización. Indica que los datos se deben conservar al menos durante diez años a partir de la finalización de un proyecto, pero no obliga a depositarlos en ningún sitio. Finalmente y desde el ámbito americano destaca el *National Institutes of Health* como ejemplo en la planificación de la preservación de datos de investigación llevado a cabo por una institución nacional. En 2008 dispuso de un mandato de investigación que obligaba a sus investigadores a depositar sus documentos peer-reviewed en *Pubmed Central* y en formatos predefinidos.

2.5. Un nuevo reto para las bibliotecas

Durante largo tiempo el cometido de las bibliotecas se ha centrado en la adquisición y preservación de colecciones documentales representativas de nuestra cultura y ciencia. En cierta manera se había asumido que la adquisición de la publicación de la investigación científica terminaba en los registros bibliográficos y en las publicaciones finales. En estos momentos el nuevo reto de las bibliotecas recae en el problema que plantea la preservación de la memoria digital, y de entre los activos que la componen, son los datos de investigación los que representan el mayor de los desafíos.

Una de las características de la ciencia moderna es la importancia creciente que se da a los datos de investigación, costosos y difíciles, si no imposibles, de repetir. Bryan Heidorn (2011) apunta que un argumento para que las bibliotecas tomen las riendas en *data curation* lo defiende el hecho que las misiones de la mayoría de las bibliotecas universitarias centran su objetivo en conseguir una mayor calidad del conocimiento, el aprendizaje y la investigación mediante la adquisición y disseminación de la información. Por tanto, el papel de la biblioteca como garante de los datos de investigación queda perfectamente encajado en su cometido.

A pesar de que en Europa no existen ni siquiera en la actualidad muchos centros de datos e instituciones que ofrezcan este tipo de servicios: la *University Edinburgh* fue la primera en tener una biblioteca de datos, y en la actualidad también ofrecen servicios de soporte a la preservación de datos la *London School of Economics* y la *University of Oxford*. En Estados Unidos, Canadá y Australia, la situación es bastante diferente. En el artículo de Martínez- Uribe y MacDonald (2008) encontramos algunos ejemplos, como el de la *Data Liberation Initiative* del Canadá, en la que actualmente participan más de 60 bibliotecas y que pretende facilitar a todos sus miembros servicios de información e infraestructuras que aseguren la efectividad de los servicios de datos de cada institución.

En el ámbito anglosajón es donde las bibliotecas académicas se han movilizado con más rapidez para empezar a afrontar el nuevo reto del *data curation*. Como hemos visto, en Estados Unidos, algunas universidades participan activamente en la iniciativa de la *National Science Foundation* (NSF) para establecer una gestión de los datos de investigación, mientras en Europa y el Reino Unido el liderazgo recae en el DCC. La *Cornell University*, la *John Hopkins University*, el *Massachusetts Institute of Technology*, *University of Illinois* y la de *University of New Mexico* son algunas de las instituciones que en la actualidad ya han desarrollado y puesto en marcha servicios para la gestión de los datos de investigación.

Las bibliotecas académicas deben acoger la responsabilidad de continuar siendo las depositarias naturales de la producción de la sociedad del conocimiento. Para llevar a cabo este cometido es necesario que las organizaciones, las bibliotecas, preserven el contenido digital (el objeto), pero también el software y el hardware. Para Inge Angevaare (2009) la preservación de los datos a largo plazo requiere inversión de capital pero también experiencia. Y es justamente en este punto donde pueden entrar en acción las bibliotecas universitarias y de otros centros de investigación.

2.6. Bibliotecas en acción: algunos ejemplos

Algunas bibliotecas han sido pioneras en el desarrollo de políticas, acuerdos con investigadores y servicios en *data curation*. La University of Edinburgh fue la primera biblioteca de datos en el Reino Unido, y en la actualidad la *London School of Economics* y la *Oxford University* se han sumado a ofrecer servicios de este tipo. La *Cornell University* es una institución que en 2005 empezó a colaborar con la agencia estatal de financiación de la investigación americana, la *National Science Foundation* (NSF) en un primer proyecto de *data curation*. Posteriormente impulsó un repositorio de datos llamado DataStaR con el objetivo de apoyar la colaboración y el *data sharing* y para ser un modelo para futuras iniciativas de otras instituciones de investigación.

Otro de los proyectos llevado a cabo por la biblioteca de la *Cornell University* fue la creación en el 2006 de un grupo de trabajo enfocado a intercambiar información sobre *data curation*. Fruto de este grupo, se redactaron algunas recomendaciones como la colaboración de la universidad con los productores de datos, la consolidación de un grupo de trabajo de *data curation* interuniversitario en el que la *Cornell University* ocuparía el rol de líder, y la propuesta de creación de nuevos servicios. Estos servicios incluyen: asistencia a los investigadores en la creación de planes de gestión de datos; formulación de buenas prácticas e información sobre la gestión de datos; formación de los investigadores en temas de propiedad intelectual y temas relacionados con datos; desarrollo de políticas institucionales de archivo de datos mediante colaboración, y sinergias entre la *Cornell University Library* y otras universidades.

La *Purdue University Library* se distingue por crear en el año 2006 y en colaboración con profesorado del departamento de informática y otras disciplinas, el *Distributed Data Curation Center* (D2C2). Se trata de un proyecto liderado por los bibliotecarios temáticos de esta universidad para poner en marcha actividades de apoyo a la gestión de datos de investigación. Una de las actividades que empezaron a desarrollar fue su especialización en la realización de entrevistas sistemáticas

con los investigadores para comprender mejor su proceso de creación y de las tipología de datos. Así mismo, llevaron a cabo un programa de investigación sobre las causas de presencia o ausencia de *data sharing* entre los investigadores y en 2009 lanzaron uno de los primeros repositorios de datos, el e-Data.

El reconocido *Massachusetts Institute of Technology* (MIT) también fue uno de los centros pioneros en la creación de un grupo de trabajo sobre datos (*Data Initiatives Group*) en 2006, que tenía como finalidad la elaboración de entrevistas con los grupos de investigación para compartir información sobre el desarrollo de los servicios de datos. El resultado se transformó en una página web especializada dentro del portal de la Biblioteca del MIT. Se desarrolló un Dspace i SIMILE, un proyecto para el desarrollo de herramientas de código abierto enfocado a la mejora de la interoperabilidad entre los objetos digitales y metadatos, y desde el 2009 trabajan en el análisis del papel de los repositorios institucionales para la gestión de los datos de investigación.

La biblioteca de la *University of Cambridge* fue la primera en el Reino Unido en crear un repositorio institucional con una atención especial a los datos de investigación además de las publicaciones electrónicas.

2.7. El papel del bibliotecario

Aunque las iniciativas de *data curation* son relativamente recientes, los bibliotecarios hace tiempo que están acostumbrados a tratar con datos, y han ido evolucionando al ritmo de los avances tecnológicos. Los antecedentes en la preservación de datos de investigación en el ámbito de las bibliotecas se remontan a la creación de repositorios institucionales. El establecimiento de repositorios de confianza en muchas de las bibliotecas universitarias conlleva, tal y como afirma Walters (2009), a la definición del ciclo de vida de estos documentos digitales de investigación, y facilita a los bibliotecarios su conversión hacia *data curators* o *data librarians*.

Los *data librarians* trabajan mano a mano con los investigadores, asesorando sobre sistemas de búsqueda de datos o formas de acceso, y formando sobre la utilización de los datos. Esto puede llevar al profesional de la información a formar parte integral de equipos de investigación (Martinez-Urbe y Macdonald, 2008).

El desarrollo del *data curation* presenta un panorama de retos y oportunidades para el profesional de la información. Para muchos autores plantea un cambio profundo en los roles, competencias y la formación de los bibliotecarios interesados en trabajar este ámbito. Martin Lewis (2010) consideraba que puesto que los datos de investigación eran parte integral de la creación de conocimiento, su gestión resultaba una extensión natural de la tarea de los bibliotecarios en su afán por proveer acceso a las publicaciones científicas. En esta línea, el desarrollo de una gestión de los datos de investigación puede ser una oportunidad para las bibliotecas universitarias de acercarse a sus investigadores y a otras instituciones o agencias de la universidad. Lewis sugiere nueve líneas estratégicas planteadas de forma piramidal, relacionadas con diversas áreas de trabajo con las que los bibliotecarios ya están familiarizados, para que las bibliotecas se comprometan con la gestión de los datos de investigación:

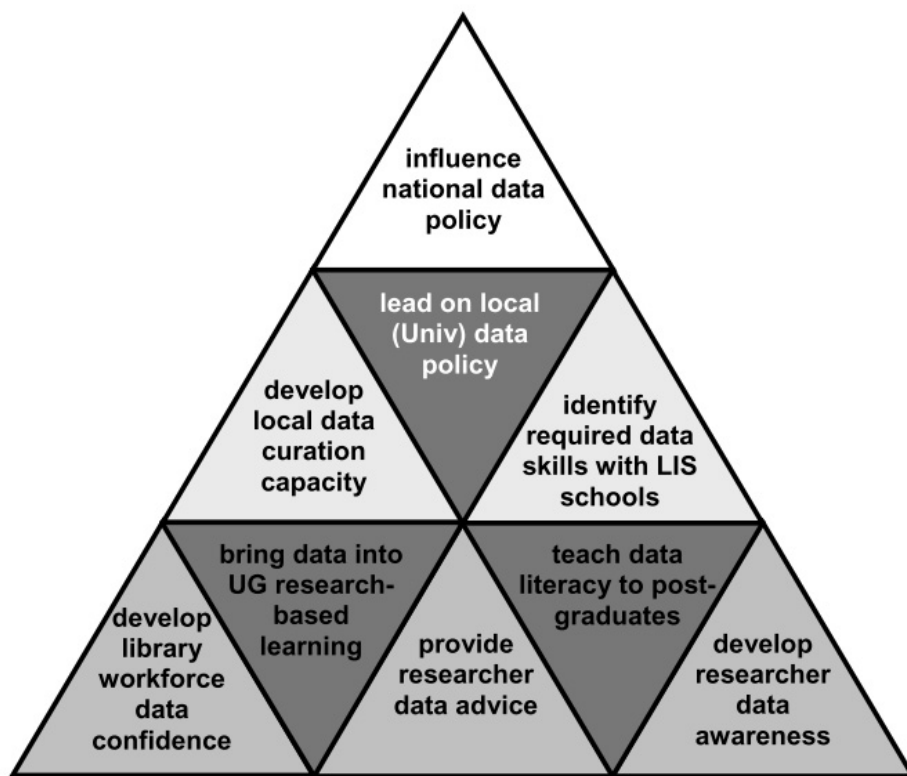


Figura 4. La pirámide de Lewis (2010)

Estas áreas de actividad son:

- Desarrollar la confianza de la biblioteca sobre los datos: se basa en la importancia de aumentar el nivel de conocimiento de la biblioteca en relación con la investigación (*e-research*) y las distintas opciones de gestión de datos. Para ello se propone mantener conversaciones con los distintos equipos académicos y de investigación sobre el tema de los datos, así como estar permanentemente actualizados sobre las principales políticas de financiación, el acceso abierto y datos de la universidad y otras instituciones nacionales.
- Fomentar un uso ético de los datos: se trata de poder resolver las dudas que la mayoría de investigadores tienen en relación con la gestión de los datos, permisos de publicación de datos, miedos a malos usos o interpretaciones incorrectas.

- Formación sobre datos en los estudiantes de postgrado: la formación de usuarios es uno de los roles de la biblioteca. Contribuir a la formación de los futuros investigadores puede resultar clave para la futura planificación de la información.
- Acercar los datos a los estudiantes de grado: en la misma línea que el punto anterior, se propone familiarizar el uso y tratamiento de los datos a usuarios más iniciales.
- Formar en capacidades para el *data curation*: mediante el uso de guías, herramientas o formaciones con grupos reducidos de investigadores
- Identificar las habilidades necesarias juntamente con los centros de formación en documentación: los directores de la biblioteca deben reconocer las necesidades que tiene su personal y potenciar el trabajo conjunto con las instituciones de formación de biblioteconomía y documentación.
- Liderar las políticas locales de datos: la biblioteca y su dirección debe entender los retos y beneficios de la gestión de datos y trabajar conjuntamente con la institución como líder.
- Influenciar las políticas nacionales de datos: las bibliotecas deben ser activas en el desarrollo de políticas nacionales sobre los datos de investigación.

El papel del bibliotecario de datos (*data librarian*) debe ser en definitiva el de trabajar muy estrechamente con los investigadores, asesorando, formando y asistiendo en la utilización y gestión de los datos. Tal y como apunta Martínez Uribe y Macdonald (2008), la familiarización de los bibliotecarios universitarios con el mundo de los repositorios académicos y del acceso abierto, facilita la adaptación a sus nuevas tareas de *data librarian* y les permitirá jugar un papel importante en la gestión y preservación de los datos de investigación.

Una cuestión que resulta necesaria abordar al hablar del papel del bibliotecario es el de las competencias que estos profesionales deben reunir para poder llevar a

cabo su cometido en el terreno del *data curation*. Existen en la actualidad una gran variedad de artículos, estudios y encuestas que se han centrado en estudiar este punto (Tenopir, y otros, 2012). De forma general, se insiste en competencias sobre la recolección, descripción y procesamiento de los datos de investigación, y poco a poco estas necesidades se van reflejando, e en adaptaciones de los programas de estudios y formaciones curriculares de los bibliotecarios y documentalistas. En este sentido destacan los programas de formación de instituciones académicas americanas como la University of Arizona, que ofrece la posibilidad de cursar de forma remota un grado específico en *Graduate Certificate in Digital Information Management*; la University of California at Berkeley, dispensa un *Master of Information Management and Systems*; la University of Illinois at Urbana-Champaign ha desarrollado un programa específico sobre data curation, el *Data Curation Education Program* (DCEP) (Keralis, 2012). Fuera del ámbito estrictamente universitario también han surgido proyectos dirigidos al *data curation*, como el *Digital Curator Vocational Education Europe* (DigCurv), impulsado por la Comisión Europea y que pretende establecer un marco curricular para la formación profesional en *data curation*.

Una de las propuestas más interesantes sobre las habilidades que a nivel genérico deberían adquirir los profesionales de la información, es el modelo híbrido que propone Corall (2008): la conjunción de las habilidades del *Information Technology and Media* (especialista tecnológico), el *Library and Information Science* (especialista de contenidos) y el *Academic and Profesional Disciplines* (el rol más académico).

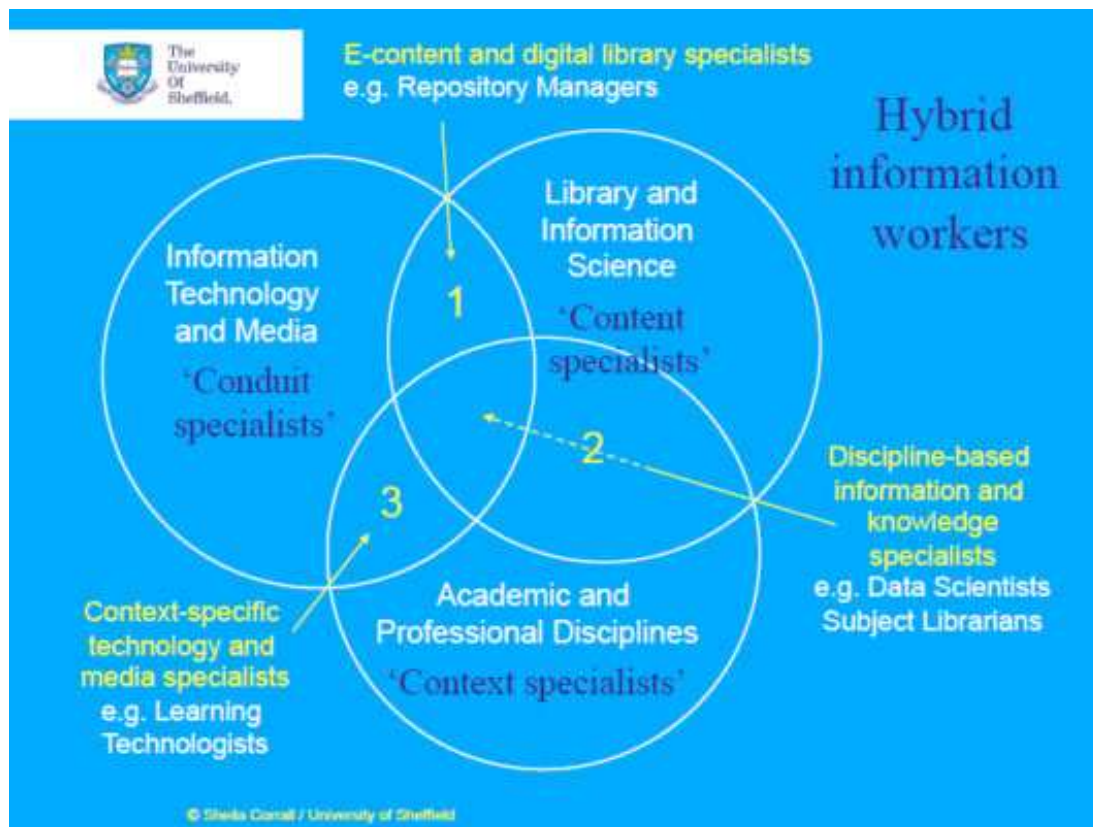


Figura 5. Perfil híbrido del profesional de la información (Corall 2008).

No existe un consenso general que permita establecer las habilidades y la formación necesaria para el desarrollo de servicios de preservación y conservación de datos de investigación. Para otros autores el papel de los bibliotecarios varía y depende de su contexto específico, de sus intereses y de las políticas de la biblioteca o institución en la que desarrolle su labor, de la apertura y flexibilidad de los investigadores con los que trabaja y de los recursos de los que se dispongan (Reilly 2012). Siguiendo este discurso, la propuesta de Donnelly (2008) se basa en cuatro roles diferenciales que pueden estar en contacto con el *data curation*, y que quedan bien definidos en la figura 6.

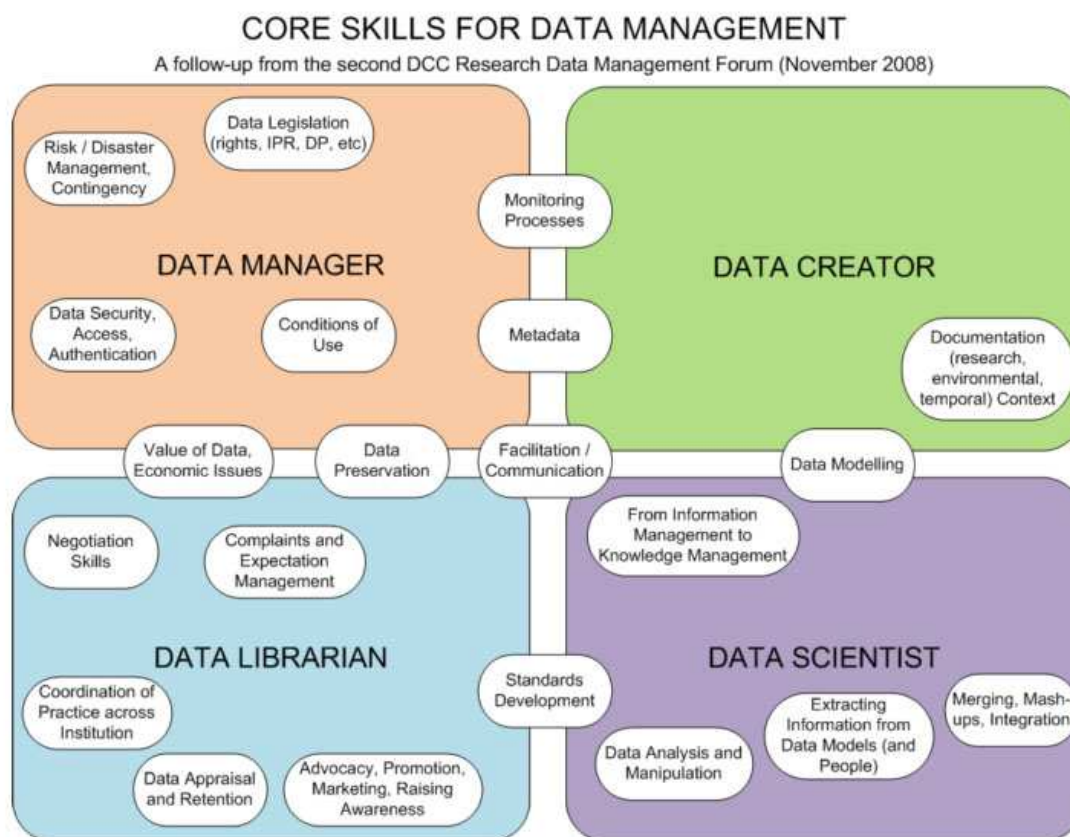


Figura 6. Core Skills for Data Management, de Donnelly (2008)

Según Reilly (2012) las habilidades prioritarias que son necesarias desarrollar en los bibliotecarios que den apoyo a la gestión de los datos de investigación, son las capacidades TIC, la preservación y archivo de datos, la experiencia en temas específicos de investigación, la capacidad de evaluar datos de investigación y la capacidad de visualizar conjuntos de datos.

La principal conclusión sobre el papel del bibliotecario en relación con el desarrollo de actividades y servicios de *data curation*, es que hoy por hoy aún resulta necesaria una importante inversión de recursos para formar a los bibliotecarios interesados en este ámbito y que las bibliotecas se encuentran delante de una nueva oportunidad que requiere de cambios, inversiones y formación, y es posible que necesite incorporar o formar un nuevo perfil de bibliotecario.

3. Herramientas y recursos para la preservación de datos de investigación

A continuación desarrollamos y explicamos una breve selección de recursos y herramientas que han sido concebidas como ayudas y materiales de apoyo para el desarrollo de actividades de preservación de datos de investigación, relevantes para cualquier institución o persona interesada en desarrollar o aproximarse a un modelo de *data curation*. Las principales características de dichos recursos se han recogido en una tabla para visualizar mejor sus diferencias y actividades.

Tabla 1. Recursos y herramientas para la gestión de datos de investigación

RECURSO/ HERRAMIENTA	DESTINATARIO	MATERIALES	SERVICIOS	DIFUSION/FORMA CIÓN
Data Curation Profiles Toolkit http://datacuratiprofiles.org/	- Bibliotecarios - Investigadores - Instituciones (a modo general)	-Guía de usuario - Manual de entrevista - Plantilla de trabajo - Plantilla de perfil		- Foro de la comunidad - Simposios - Workshops
MIT Libraries: Data Management and Publishing http://libraries.mit.edu/guides/subjects/data-management/plans.html	Investigadores del MIT = personal interno	-Guía de gestión de datos (planes de gestión de datos, documentación/metadatos, formatos, organización de datos, seguridad, citación, integridad de datos, requerimientos de agencias, legalidad, compartición y archivo de datos)	Repositorio institucional Atención personalizada al investigador	Workshops
ICPSR http://www.icpsr.umich.edu/icpsr	Bibliotecarios Investigadores Profesorado	Guía para un plan de gestión de datos efectivo	Archivo de datos Formación	Summer program Programa de difusión

rweb/landing.jsp	Estudiantes		a medida	
Digital Curation Exchange http://digitalcurationexchange.org/	Bibliotecarios Investigadores Profesorado Estudiantes	Materiales de aprendizaje, Materiales para la práctica Materiales para la investigación Materiales para la educación	Listado de recursos Preguntas y respuestas Materiales	Foro Grupos de trabajo Bolsa de trabajo Eventos
DCC http://www.dcc.ac.uk/resources	Bibliotecarios Investigadores	Manuales de referencia Manual del ciclo de datos Planes de gestión de datos Manuales y herramientas para políticas de datos Guía en cuestiones legales Estudios de casos Estándares	Revista especializada Servicios de formación y prácticas, jornadas	Jornadas Eventos Comunidad

Data Curation Profiles Toolkit



Se trata de un proyecto desarrollado por iniciativa de dos instituciones americanas, la *Purdue University* y la *Illinois University*. Esta herramienta, pensada tanto para profesionales de la información (bibliotecarios y documentalistas), cómo para los científicos creadores de los datos, pretende ofrecer pautas y requerimientos para la creación y preservación de los datos de investigación. Esta iniciativa desarrolla distintos perfiles temáticos sobre los que elabora campos y plantillas genéricas, con cierto nivel de estandarización, para que la gestión y preservación de los datos de investigación sea lo más flexible y aplicable posible a cualquier disciplina.

Para obtener la creación de un “perfil” se basan en entrevistas, publicaciones, materiales y documentación específica de cada campo de investigación. En este recurso resulta de especial interés los materiales que pone a disposición de todos los usuarios: una guía de usuario, en la que se explican cada uno de los pasos de cómo usar esta propuesta metodológica para la preservación de datos de investigación; un manual de entrevistas, que ofrece pautas, recomendaciones, reflexiones y preguntas concretas e instrucciones para el entrevistador para afrontar las entrevistas; una plantilla de trabajo, un documento estándar realizado para que el investigador rellene toda la información sobre datos que le será preguntada durante la entrevista; una plantilla de perfil, se trata de una plantilla predeterminada con todos los campos que formarán el nuevo perfil de datos que se cree.

Su nivel de aplicación es a tres niveles:

- a nivel individual facilita las bases para el desarrollo de entrevistas sobre datos entre bibliotecarios y grupos de investigación y proporciona herramientas a los grupos de investigación para reflexionar sobre el uso de los datos
- a nivel institucional proporciona documentación para la gestión de datos y además los distintos perfiles pueden ser usados para identificar los tipos de herramientas, estructuras y responsabilidades para posibles servicios de datos.

De forma genérica, *Data Curation Profiles Toolkit* se puede concebir como una guía para desarrollar servicios de datos en cualquier institución, para mejorar la comprensión de las tipologías de datos y conocer las necesidades de los investigadores.

En la actualidad se encuentran disponibles un total de 27 perfiles de datos de investigación distintos, y aunque siguen siendo mayoría los campos de las ciencias, encontramos algunos ejemplos de ciencias sociales y humanidades como los perfiles de lingüística y sociología.

MIT Libraries: Data Management and Publishing



Las bibliotecas del *Massachusetts Institute of Technology* han sido pioneras en la preocupación por la gestión y preservación de datos de investigación. Escogemos el recurso desarrollado por las bibliotecas de esta institución que tiene como objetivo dotar de servicios y ayuda a su comunidad investigadora en relación con los datos.

En su espacio web podemos encontrar distintos materiales destinados básicamente a los investigadores, entre los que destacan materiales como una guía para la gestión de datos, con recomendaciones acerca de: planificación, organización y formato de los datos, seguridad, temas legales y el archivo de datos. De entre sus servicios personalizados resulta interesante la atención y ayuda especial que ofrecen para solucionar las dudas procedentes de los requisitos de las agencias de financiación americanas.

Así mismo, en tanto que alta institución académica, la biblioteca del MIT pone al servicio de su comunidad la posibilidad de almacenar los datos de investigación en un repositorio, o bien recomienda otros repositorios de datos adecuados a cada campo de investigación.

Inter-university Consortium for Political and Social Research (ICPSR)



Desde hace más de diez años, este consorcio americano con sede en la *University of Michigan*, pone a disposición de toda la comunidad de investigación en ciencias sociales un servicio de acompañamiento, apoyo y formación en relación a los datos

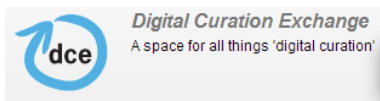
de investigación. Mediante el desarrollo de formaciones sobre el acceso de datos y la preservación, tiene como objetivo la máxima difusión de las ciencias sociales.

Constituye un instrumento muy amplio y variado que puede resultar de gran utilidad para un público de niveles muy heterogéneos, interesado en los datos de investigación. Aunque es necesario afiliarse a nivel institucional al consorcio, que cuenta con más de setecientos afiliados, para acceder a servicios específicos, también dispone de un conjunto de herramientas y materiales relevantes de libre acceso para las actividades de *data curation* que nos ocupan. Por un lado, dispone del repositorio de datos de investigación en ciencias sociales más grande del mundo, en el que es posible navegar por distintas facetas y acceder a los sets de datos.

Ofrece servicios personalizados de *data curation*, almacenamiento a largo plazo, difusión mundial y servicio personalizado de atención al usuario. Así mismo pone a disposición de cualquier usuario la consulta de materiales y guías sobre la creación y preservación de datos, dirigidos tanto a profesionales como estudiantes (vídeos, ejercicios y recursos externos). Ha desarrollado una *Guía para un plan de gestión de datos efectivo*, en la que define los elementos de un plan de este tipo, el contexto para su creación, un plan para el almacenaje de datos, ejemplos y preguntas frecuentes.

De igual manera organiza cada verano y anualmente un programa de actividades de formación muy amplio, el *Summer Program in Quantitative Methods of Social Research*.

Digital Curation Exchange



Este recurso web tiene como principal objetivo convertirse en un espacio de participación online para todas aquellas personas que de un modo u otro se vean implicadas en algún proceso relacionado con la preservación de datos de

investigación. Sus destinatarios potenciales son investigadores, profesores, estudiantes y bibliotecarios.

Esta herramienta se concibe como un foro de encuentro, en el que proponer y encontrar preguntas y respuestas, compartir ejemplos de buenas prácticas de *data curation*, así como para estar al día de eventos interesantes, nuevas herramientas e incluso ofertas de trabajo.

Está organizado sobre cuatro ejes principales: aprendizaje, práctica, investigación y educación, y cada uno de ellos ofrece herramientas, recursos, ofertas de trabajo y foros especializados. Se trata de un recurso para la formación continua de aquellas personas ya implicadas en actividades de preservación de datos de investigación, de un buen centro de recursos e intercambio de información.

Digital Curation Center Resources



Desde su creación en 2004 ha desarrollado un interesante bloque especialmente dedicado a los recursos y herramientas para la preservación de datos de investigación. Este espacio es posiblemente uno de los más completos para el profesional de la documentación interesado en el *data curation*, ya que facilita información relativa a casi todos los campos y fases en las que esta actividad está implicada: manuales, tanto básicos y de referencia para el digital curation como para la gestión del ciclo de los datos; herramientas específicas para bibliotecarios y para investigadores; acceso a revistas de cabecera (como el *International Journal of Digital Curation*). A nivel de políticas de preservación de datos dispone de una amplia y valiosa documentación, desde un análisis de las políticas de las agencias de financiación, recomendaciones sobre aspectos legales y hasta ejemplos de políticas de datos de distintas instituciones y un sumario exhaustivo de los pasos para la creación de un documento de este tipo.

El DCC propone un amplio elenco de herramientas, información de estándares, repositorios y publicaciones. Su completo programa de formación de profesionales, que combina con jornadas y eventos gratuitos, presenciales y online, programados o bajo demanda tanto para investigadores como para bibliotecarios.

4. Informe de la situación del *data curation* en la Universitat Oberta de Catalunya

El cuarto apartado de este trabajo parte de este trabajo aborda el caso de la Universitat Oberta de Catalunya (UOC) en lo que a *data curation* se refiere. Se trata de desarrollar algunos de los requisitos que dicha institución tendría que plantearse para llevar a cabo algún tipo de plan de preservación de datos de investigación.

El informe se desglosa en tres bloques:

1. En el primer apartado se realiza una introducción sobre la UOC: qué tipo de institución es, cómo está internamente organizada y quiénes son los principales implicados con la actividad de preservación de datos de investigación. Se trata de presentar la especificidad y la situación de la investigación en esta universidad.
2. En el segundo apartado se centra en la situación de los datos de investigación en la UOC, la normativa, los tipos de datos, la tecnología que se utiliza, el papel de los agentes implicados, las experiencias más próximas de *data curation*, y finalmente, se realiza un análisis DAFO para valorar todas las características analizadas.
3. El tercer apartado tratará de establecer una propuesta de actividades de actuación básicas para la mejora de la situación de los datos y la actividad de *data curation* en la UOC.

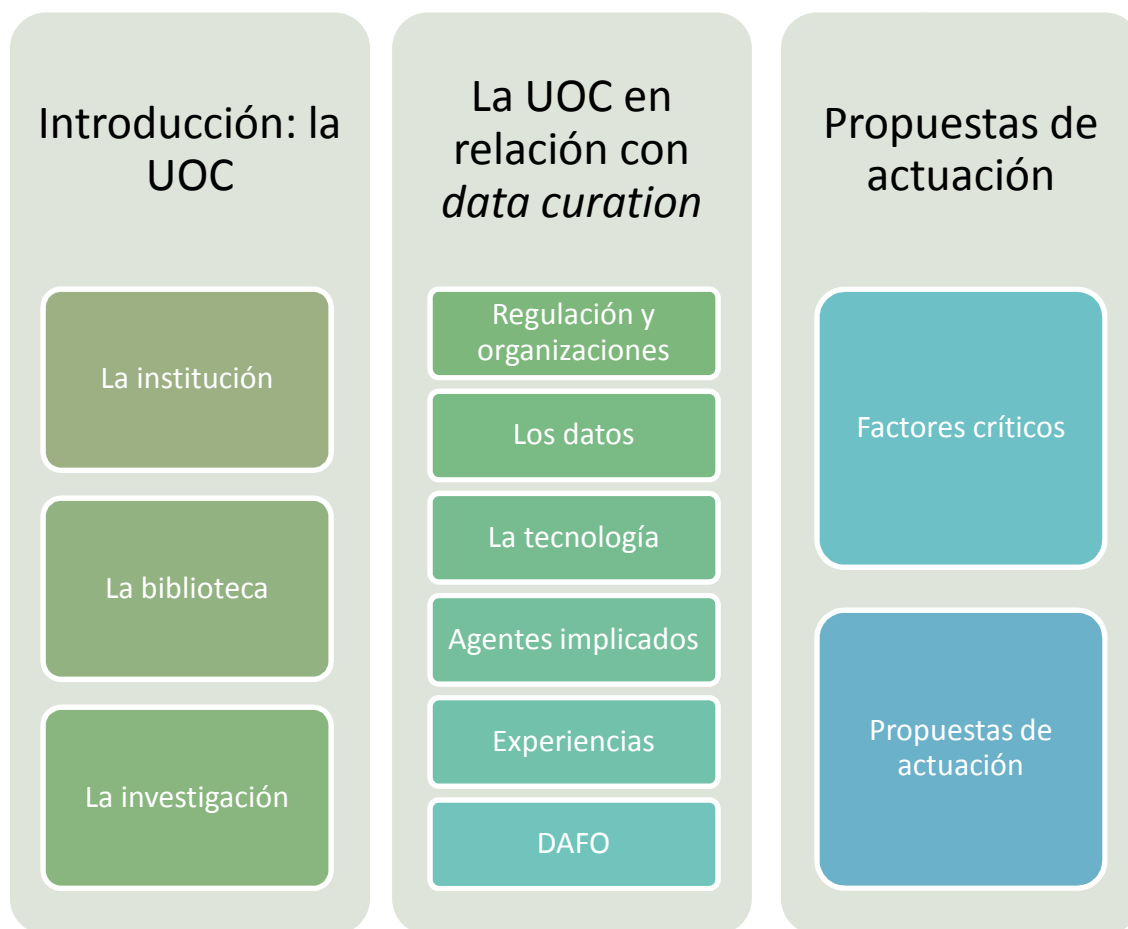


Figura 7. Estructura del informe para la UOC

4.1. Introducción: la Universitat Oberta de Catalunya

Para realizar este apartado se han consultado las informaciones disponibles en la página web de la universidad, las distintas secciones de su campus virtual, y documentación interna de la propia biblioteca virtual. Destacamos además la información facilitada por el reciente artículo de un grupo de bibliotecarias, titulado *“Bibliotecarios de investigación en un entorno virtual: nuevas habilidades, nuevos servicios”* que por un lado revisa la evolución de la Biblioteca de la UOC, y por otro, propone nuevos servicios y formación ofrecer un mejor apoyo a sus investigadores (Riera Quintero, Padrós Cuxart, y Zuñiga Ruiz, 2012).

4.1.1. La institución

En 1995 tuvo lugar la fundación de la Universitat Oberta de Catalunya y un año más tarde se inauguró el primer curso académico, ofreciendo dos cursos piloto de dos titulaciones, Ciencias Empresariales y Psicopedagogía, con 200 estudiantes y 14 asignaturas. Desde entonces, la primera universidad virtual del estado ha ido creciendo progresivamente. La oferta formativa alcanza a 15 grados, 24 másteres y posgrados y dos programas de doctorado y en el 2012 ya son casi 40.000 el número de estudiantes matriculados en esta universidad virtual.

En la actualidad, esta institución académica pionera en *el e-learning* está formada por siete áreas de estudios (Artes y Humanidades, Ciencias de la Salud, Ciencias de la Información y de la Comunicación, Derecho y Ciencia Política, Economía y Empresa, Informática, Multimedia y Telecomunicación y Psicología y Ciencias de la Educación); tres escuelas (Business School, Escuela de Cooperación, Escuela de lenguas); un Instituto Internacional de Posgrado, dos Institutos de investigación (Internet Interdisciplinary Institute o IN3, eLearn Center o eLC); y tres cátedras (Cátedra UNESCO de E-learning, Cátedra UNESCO - Fundación FCB – UOC y Cátedra de Multilingüismo Linguamón-UOC).

La UOC se caracteriza por ser la primera universidad de educación a distancia en línea del Estado español, y por tratarse de una institución pionera en la investigación de la aplicación de las nuevas tecnologías educativas en las aulas. En la actualidad es considerada un referente en los ámbitos de estudio de la sociedad de la información y del aprendizaje virtual. La particularidad de la docencia reside en que se realiza en su totalidad a través de un Campus Virtual, de forma que todos los servicios y actividades se desarrollan exclusivamente de forma virtual.

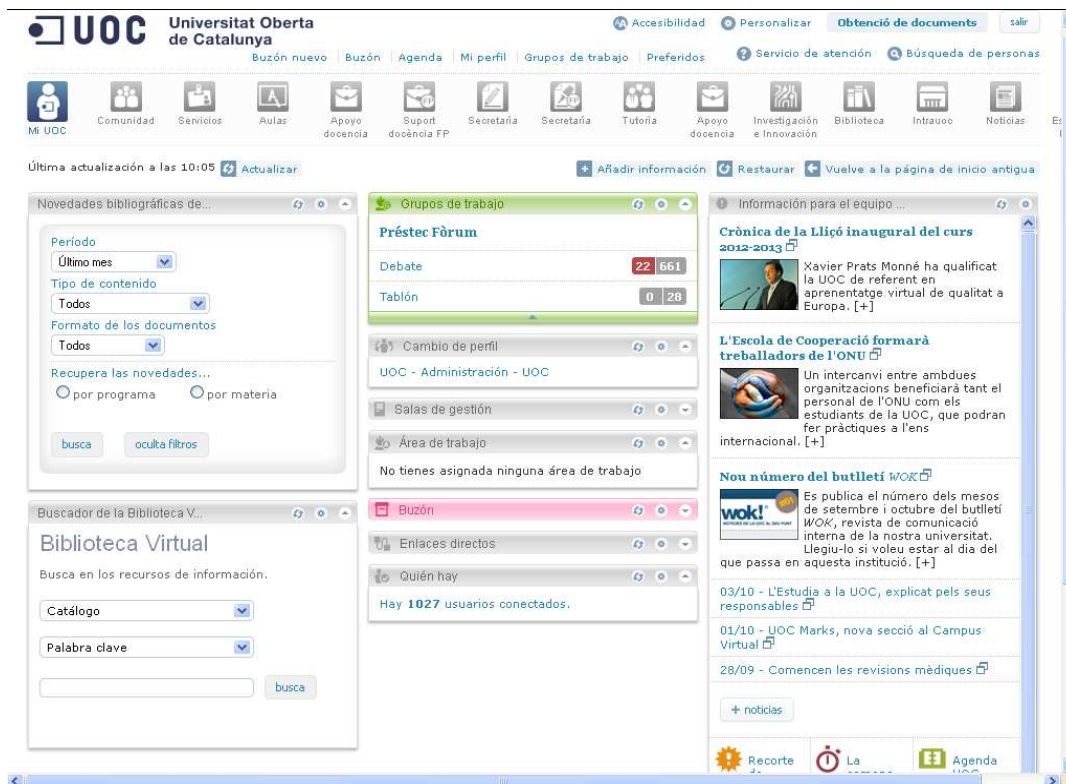


Figura 8. Campus Virtual UOC. Página de inicio

El objetivo de la UOC queda reflejado en la definición de su misión como universidad:

La misión de la UOC (Universitat Oberta de Catalunya) es proporcionar a las personas un aprendizaje duradero y oportunidades educativas. El objetivo es ayudar a las personas a cubrir sus necesidades de aprendizaje y proporcionarles un acceso completo al conocimiento, más allá y por encima de la planificación habitual y las restricciones de espacio. (Generalitat de Catalunya, 1995.)

A nivel de organización interna, la UOC se rige por dos tipos de gobierno: los órganos de gobierno unipersonales, formados por la Rectora y los vicerrectores, el secretario general y el gerente; y los órganos de gobierno colegiados, que incluyen el Patronato de la Fundación para la Universitat Oberta de Catalunya (FUOC), la Comisión Permanente de la FUOC, el Consejo de la FUOC, el Consejo de Gobierno y

la Comisión Estratégica.

Nos interesa aquí remarcar la constitución del Consejo de Gobierno, formado por la Rectora, el Gerente, los Vicerrectores (de Posgrado y Formación Continua, de Innovación e Investigación, de Tecnología, de Ordenación Académica y Profesorado), el Representante del profesorado, el Representante de los estudiantes y el Representante del personal de gestión y sobretodo conocer el Comité de Dirección Ejecutiva, el encargado de tomar decisiones sobre proyectos como el de *data curation*.

4.1.2. La Biblioteca

En 1995, el mismo año de la fundación de la UOC, se creó la Biblioteca virtual con el objetivo de ofrecer apoyo a la comunidad docente, de investigación y de gestión. Los servicios de la Biblioteca virtual de la UOC se prestan exclusivamente de forma virtual, a través del campus online. Aprovechando las posibilidades tecnológicas del campus virtual, la Biblioteca de la UOC no solamente es consultable a través de un espacio web propio, sino que se encuentra presente en el interior del aula del estudiante, ofreciendo materiales (materiales docentes de aula, módulos virtuales de asignaturas...), servicios (como el servicio de referencia virtual “La biblioteca responde”, de obtención de documentos, préstamo de libros), y acceso a recursos de suscripción (como las bases de datos o las revistas), directamente desde el lugar de interés del alumno y relacionados o personalizados con los temas del aula en la que está matriculado.

La Biblioteca ha evolucionado al mismo ritmo que la universidad, ampliando y mejorando sus servicios al usuario. Se trata de una biblioteca abierta las 24 horas del día los 365 días del año que proporciona información a sus usuarios no solo en soporte tradicional (papel) sino también en nuevos soportes como audio, vídeo, formato electrónico y Internet. Se accede a la Biblioteca virtual mediante su página web y todos los recursos están disponibles desde esta plataforma. La política de la institución se basa en adquirir recursos primordialmente en formato electrónico y

solamente en el caso que no esté disponible en este formato se comprará en papel. Si un documento no se encuentra disponible online, se puede solicitar su préstamo a domicilio, servicio que se ofrece a través de una empresa de mensajería, o bien es posible recogerlo en alguna de las numerosas sedes territoriales que la UOC tiene por toda España.



Figura 9. Recursos de la Biblioteca Virtual UOC

En la actualidad la Biblioteca virtual pone a disposición de los usuarios de la UOC un amplio abanico de servicios y recursos de información, desde el tradicional catálogo y préstamos de documentos de la Biblioteca, préstamo de lectores *e-reader*, pasando por más de 70 bases de datos de suscripción, revistas electrónicas, servicios de obtención de documentos, un repositorio institucional (O2), un espacio de libros electrónicos y hasta la implementación de una nueva herramienta *discovery tool* llamada Summon que incluye la totalidad de los recursos de la Biblioteca.

A partir de 2011 tuvo lugar un cambio organizativo interno que modificó la antigua estructura de la Biblioteca virtual, organizada en tres grupos tradicionales (servicios, recursos electrónicos y tecnología) por la creación de dos grupos

operativos principales, uno de ellos orientado a los Servicios de Biblioteca de Apoyo al Aprendizaje (SBA) y otro dirigido a los Servicios de Apoyo a la Investigación (SBR). Este cambio ha permitido desplegar una carta de servicios específica para investigadores. El grupo de SBR está formado por un conjunto de ocho bibliotecarios o documentalistas, cuatro de ellos con competencias específicas en apoyo a la investigación, como el conocimiento de herramientas bibliométricas, gestión de repositorios y elaboración de bibliografías, tres de ellos que gestionan los servicios de la biblioteca y una directora del grupo, que han desplegado una carta de servicios de apoyo a la investigación (Universitat Oberta de Catalunya, 2010).

La Biblioteca virtual de la UOC es uno de los agentes implicados de forma transversal durante todo el proceso de investigación, desde el nacimiento de una idea o propuesta, pasando por la fase de recogida y análisis de datos hasta la difusión y la posterior evaluación en los procesos de acreditación de su calidad (Riera Quintero, Padrós Cuxart, y Zuñiga Ruiz, 2012).

4.1.3. La actividad investigadora

La actividad de investigación de la UOC se desarrolla principalmente por el profesorado de la universidad y el personal de investigación. Los treinta y seis grupos de investigación que existen en la actualidad (catorce de ellos reconocidos por la Generalitat de Catalunya en la convocatoria de 2009) reúnen un total aproximado de trescientos investigadores vinculados a una área de estudio (de los siete que hemos repasado en el apartado anterior) o bien pertenecen a uno de los dos centros de investigación de la universidad: el IN3, con estudios multidisciplinares de la sociedad de la información y el conocimiento; y el *eLearnCenter*, centro de investigación, innovación y formación en el aprendizaje virtual dedicado esencialmente a la educación superior y a la formación a lo largo de la vida.

Se potencia el trabajo transversal para lograr una mejor colaboración de todos los agentes implicados en el apoyo a la investigación. Destaca la prominencia de las

ciencias sociales, un aspecto que resulta relevante cuando tratamos con temáticas de *data curation*. Gracias a estos propósitos podemos consultar un mapa del conocimiento de la UOC, que ofrece una visión global del conocimiento generado en la institución mediante la identificación de los principales grupos de investigación, las líneas de R+D+i y sus respectivos ámbitos temáticos:

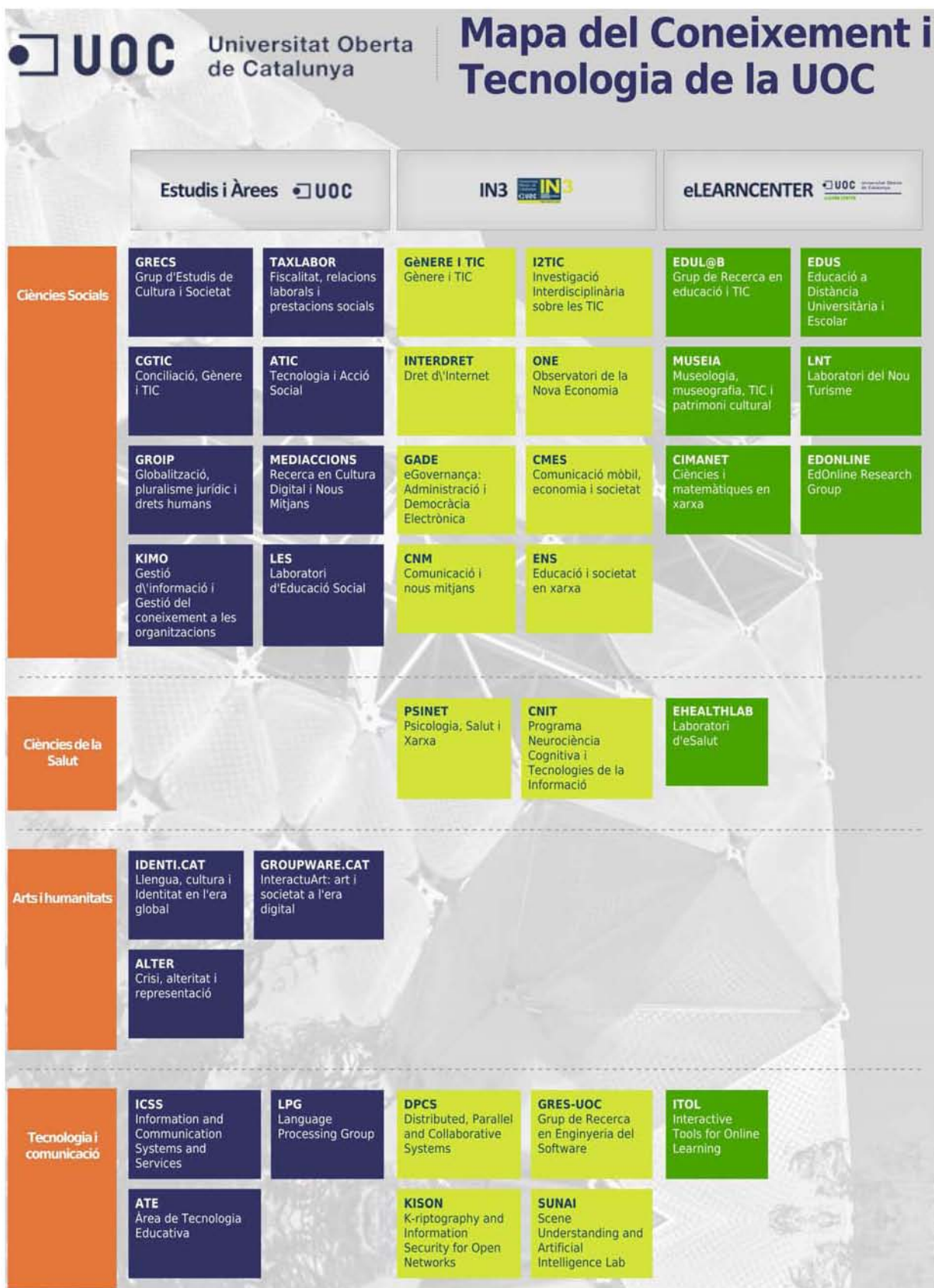


Figura 10. Mapa de conocimiento de la UOC

4.2. *Data curation* en la UOC

4.2.1. *Comentarios previos*

A continuación se describen los contenidos que serían necesarios analizar para llevar a cabo un proyecto relacionado con la preservación de datos de investigación en el contexto de la UOC. Partimos aquí de la suposición de la existencia de una voluntad por parte del Vicerrectorado de Investigación e Innovación de encargar la realización de un estudio o informe sobre la situación de la Universidad en relación con los datos de investigación. Las conclusiones de un informe de este tipo tendrían como objetivo facilitar la posterior redacción de un Plan Director para la preservación de datos de investigación.

Los puntos a tratar van a ser los siguientes:

- a) Marco de regulación legislativo y el papel de las organizaciones.
- b) Situación de los datos de investigación en la UOC.
- c) Tecnología UOC y los datos de investigación.
- d) Agentes implicados.
- e) Experiencias previas.
- f) Análisis DAFO.
- g) Propuestas de actuación.

El planteamiento de un proyecto de gestión de datos de investigación requiere como paso inicial una reflexión sobre el marco contextual de la institución. El análisis del marco externo e interno de la UOC tiene como objetivo principal que la institución, y sobre todo las personas directamente implicadas con la gestión de los datos de investigación, tengan un marco de referencia sobre el *data curation*. Un estudio reciente de un grupo de trabajo del LIBER (Christensen-Dalsgaard, 2012) concluye que prácticamente la totalidad de las bibliotecas académicas europeas considera altamente importante ofrecer nuevos servicios relacionados con la gestión de los datos de investigación, pero al mismo tiempo son muy pocas las que

han iniciado el proceso.

4.2.3. Marco de regulación legislativo y el papel de las organizaciones

El objetivo de realizar un análisis del marco legal y organizativo tanto a nivel institucional como a nivel estatal tiene como finalidad mostrar la legislación vigente e relación al *data curation*, así como los organismos de referencia españoles y de la UOC.

En España, las iniciativas legales de carácter estatal relacionadas con la gestión y preservación de los datos de investigación se caracterizan por la falta de un referente claro. Resulta relevante que en el artículo 37 de la nueva ley de la Ciencia y la Tecnología (España, 2011), se trate el tema de la difusión en Acceso Abierto pero que no se mencione en ningún capítulo conceptos relativos a la preservación de datos de investigación.

Nos encontramos, pues, en un momento de vacío legal a nivel estatal, al tiempo que se están gestando los primeros documentos y guías de buenas prácticas, así como la creación de los primeros grupos de trabajo. Ninguna institución u organismo ha tomado, hasta la fecha, las riendas de este asunto, a diferencia de otros países como el Reino Unido, Estados Unidos o Australia. A pesar de ello podemos destacar el interés por parte del grupo Recolecta (Recolector de Ciencia Abierta) de la Fundación Española para la Ciencia y la Tecnología (FECYT), que impulsa y coordina la infraestructura nacional de repositorios científicos digitales de acceso abierto y que recientemente está promoviendo la realización de conferencias y webminars sobre *data curation*.

La misma Fecyt está colaborando con la iniciativa OpenAiRE PLUS liderada por la Comisión Europea, que tiene como objetivo principal ampliar el horizonte del acceso abierto incluyendo el depósito de los datos de investigación de los proyectos financiados en el VII Programa Marco. También desde la Fecyt, se ha creado un grupo de trabajo orientado a la creación de buenas prácticas en la gestión de los datos de investigación, y que tiene como objetivo la creación del primer documento

del estado español sobre la gestión, preservación y almacenamiento de los datos de investigación en los repositorios. A fecha de hoy (octubre del 2012) no ha emitido ningún informe.

Por otro lado también cabe destacar las recientes iniciativas de aproximación al estudio del *data curation* organizadas por GrandIR, una empresa de apoyo a los repositorios institucionales universitarios y a centros de investigación liderado por Pablo de Castro, y de la Universitat Politècnica de Catalunya. Durante la sesión técnica que tuvo lugar en la sede institucional de la UOC de Barcelona el noviembre de 2010 se comentaron los *Avances en la gestión de los datos de investigación en España* (VV.AA, 2012).

Estudios recientes apuntan a que en el próximo programa marco de investigación, el FP8, la Comisión Europea incluirá la gestión de datos de investigación para sus proyectos (Zimmermann, 2012). En lo que a organismos se refiere, existen diversas iniciativas que se han mencionado anteriormente, como: *Data Curation Center* (DCC); *DISC-UK DataShar*, que trabaja para obtener repositorios con alojamiento de grupos de datos y para potenciar el crecimiento del movimiento de datos en acceso abierto; *JISCMRD (Management Research Data)*; o el proyecto DAF (*Data Audit Framework*). Finalmente destacamos la reciente iniciativa de la organización EUDAT (*European Data Infrastructure*) que aunque más enfocada al *data sharing*, también parece empezar a preocuparse por el *data curation* y la colaboración global en las infraestructuras de datos.

En relación con el posicionamiento de la UOC, aunque en la actualidad no existe ninguna política interna respecto a la creación, uso o preservación de los datos de investigación, existe un antecedente clave que puede interpretarse como un indicador fiable del posicionamiento de la institución. Se trata del mandato institucional de la UOC (Universitat Oberta de Catalunya, 2010), que apoya al acceso abierto de la producción científica. Durante el proceso de puesta en marcha y elaboración, la biblioteca virtual lideró el proyecto mediante la redacción de la guía para profesores e investigadores.

4.2.4. Gestión de los datos de investigación en la UOC

En la actualidad, la UOC no cuenta con una normativa definida para la gestión y preservación de los datos de investigación en la UOC. No se ha realizado aún ningún tipo de informe o estudio específico sobre la tipología de los datos de investigación creados en la UOC, ni a nivel institucional ni de cada una de las áreas de conocimiento. Así mismo, no se dispone de información exhaustiva y fiable sobre la localización de los datos durante y después de los procesos de investigación ni existe ninguna política de creación, estandarización ni almacenamiento de este tipo de datos por parte de la universidad, estudios ni grupos de investigación.

Con la finalidad de conocer de forma más cercana las actividades y necesidades del personal investigador de la UOC por parte de la biblioteca, el pasado año 2011 se programaron un conjunto de reuniones y entrevistas con los principales grupos de investigación de la universidad, tanto los consolidados como los emergentes. Esta iniciativa del grupo de Biblioteca de Apoyo a la Investigación (SBR) tenía como principal objetivo establecer contactos entre la biblioteca y los investigadores, conocer mejor sus actividades académicas y de investigación para poder ajustar los servicios de la Biblioteca a sus necesidades reales. A través de dicha iniciativa es posible obtener algunos datos e impresiones sobre el uso y gestión de los datos frutos de la investigación, así como de una idea general de la tipología de datos que se generan en la UOC (entrevistas, encuestas, grabaciones, estadísticas, códigos de programación).

Al final del proyecto fueron 44 el número de investigadores que participaron en las entrevistas, pertenecientes a 11 grupos de investigación distintos (ATIC, EDUL@B, GADE, GRECS, GROIP, i2TIC, INDENTL.CAT, KIMO, KISON, MEDIACCIONS i TAXLABOR). Las preguntas del cuestionario que se propuso a los investigadores incluían cuestiones sobre el modo de gestionar la información durante y después de la investigación:

¿Cómo organizáis y compartís la información, artículos, ponencias, datos estadísticos, códigos, transcripciones de entrevistas, etc. que utilizáis durante vuestro proceso de investigación o redacción de un artículo? ¿Cómo gestionáis la documentación e información creada durante el proceso de investigación cuando el artículo ya ha sido publicado?

El resultado de estas entrevistas concluye que los datos producto de la actividad científica de la UOC tienen una gestión y procesos de almacenamiento tan dispares como investigadores. La mayoría de los investigadores guardan los archivos de datos en sus ordenadores personales, sistemas de almacenamientos particulares o en el mejor de los casos, en servidores externos, propios o depositados en la nube, compartidos con otros colegas. Se confirma una total falta de organización y guía que desde el punto de vista institucional guíe o facilite a los investigadores de normas o recomendaciones para la preservación de sus datos.

4.2.5. Tecnología UOC y los datos de investigación

A nivel tecnológico, la Universitat Oberta de Catalunya cuenta con dos herramientas relacionadas directamente con la investigación: el repositorio institucional (RI) conocido como O2 (Oberta en Obert), y el sistema CRIS, conocido como GIR.

Por un lado, se ha promulgado un mandato institucional que se apoya en el repositorio institucional O2, desarrollado con el software de código abierto Dspace. Se trata de un portal que recoge, difunde y preserva las publicaciones digitales en acceso abierto producidas por los miembros de la UOC en el desarrollo de las actividades de investigación, docencia y gestión y que fue gestionada e implementada por la misma biblioteca. Esta herramienta tiene como objetivo garantizar la documentación digital generada por la propia universidad y garantizar su accesibilidad en el futuro y potenciar el acceso abierto a la ciencia siguiendo la vía verde, esto es facilitando el autoarchivo de las publicaciones

(López-Pérez y Vaquer-Suñer, 2011).

Por otro lado la UOC dispone de un CRIS (Sistema de Gestión Integral) llamado GIR, mediante el cual se gestiona la actividad investigadora de la universidad. Esta herramienta, basada en el sistema Universitas XXI permite además la generación de CVN (Currículos Vitae Normalizados) y se encuentra plenamente integrada con el repositorio institucional de la UOC, el anteriormente mencionado O2. Esto se traduce en que en el momento en que un investigador introduce una nueva publicación en el GIR, automáticamente se genera el salto hacia el repositorio. Esta interoperabilidad permite depositar el texto de la publicación, evita duplicidades e incrementa el fondo del repositorio (Prieto, 2012).

4.2.6. Agentes implicados

Los agentes implicados en cualquier actividad de preservación de datos de investigación que se quiera desarrollar en la UOC son los siguientes: la biblioteca, los investigadores, la Oficina de Apoyo a la Investigación y la Transferencia (OSRT), el Vicerrectorado de Investigación e Innovación y Aplicaciones y procesos.

El Vicerrectorado de Investigación e Innovación, como organismo de gobierno de la universidad responsable del IN3 y del *eLearnCenter*, se preocupa por el desarrollo de la actividad científica y de liderar propuestas y proyectos para su mejora y evolución.

La OSRT trabaja para dar respuesta a las necesidades de aquellos profesionales que participan en la actividad de I+D+i de la UOC. Entre sus principales actividades destacan el apoyo a las necesidades de todas las fases de un proyecto de I+D+i, desde la búsqueda de fuentes de financiamiento hasta su ejecución, pasando por un amplio seguimiento y asesoramiento, por el ofrecimiento de herramientas, servicios y recursos a todo el personal vinculado a un proyecto de I+D+i para facilitar el desarrollo de sus actividades en la universidad.

La Biblioteca como agente directamente implicada en los procesos ya existentes de

gestión y preservación de las publicaciones de la universidad y más concretamente desde el Grupo de Apoyo a la Investigación, podría ser el órgano responsable de la coordinación de todos los agentes para la elaboración, implementación, despliegue y formación en *data curation*. A pesar de la existencia del grupo especializado SBR, y de la presencia de dos bibliotecarios formados y centrados en el repositorio institucional, no existe en la actualidad ningún bibliotecario con un rol y con habilidades específicas para desarrollar el papel de *data librarian*.

En la actualidad no existe ningún grupo de trabajo, iniciativa, manual, recomendación ni política de gestión o creación de datos de investigación. No solamente como autores de los datos, sino por la importancia que pueden tener en el proceso de motivación y concienciación de la comunidad científica. Los investigadores como agentes directamente creadores de los datos de investigación, tienen un papel clave.

El grupo de Aplicaciones y Procesos, es el grupo de especialistas orientados a asegurar la evolución y mejora de las herramientas tecnológicas de la UOC diseñando, desarrollando o adquiriendo los programas informáticos para resolver las necesidades de implementación tecnológica.

4.2.7. Experiencias externas de *data curation*

Si se busca en el entorno externo más próximo a la UOC, dentro del ámbito universitario se dispone de algunos ejemplos de aplicación de planes de gestión y preservación de datos de investigación, como por ejemplo, el de la Universitat Pompeu Fabra (UPF). La UPF es una de las pocas instituciones académicas, tanto a nivel estatal como a nivel autonómico de Catalunya, que ha implementado recientemente un plan de publicación en acceso abierto de los datos de investigación vinculados a un programa específico de lenguaje computacional. En este caso, han aprovechado el repositorio institucional de publicaciones para crear colecciones específicas de datos. Se trata de una experiencia muy próxima y un ejemplo que vale la pena detenerse a analizar y consultar (Arano y otros, 2011).

Otra experiencia a resaltar es el caso de ADDI (UPV/EHU, 2012), el Repositorio Institucional de la Universidad del País Vasco. Este repositorio basado en la plataforma Dspace, permite el almacenamiento de artículos y publicaciones científicas de la universidad desde el 2011, y está experimentando con el tratamiento de los datos de investigación generados por un grupo de investigación de la UPV/EHU en el ámbito del Análisis del patrimonio histórico-artístico.

Desde el sector editorial destacan otro tipo de iniciativas que pueden resultar muy interesantes de cara al futuro, como la impulsada por el proveedor de bases de datos *Elsevier-Science Direct*, que el 2010 anunció la puesta en marcha de un proyecto que, en colaboración con la biblioteca de datos PANGAEA (*Publishing Network for Geoscientific & Environmental Data*) pretende interconectar los datos de investigación con la producción científica. De esta manera los grupos de datos depositados en PANGAEA que correspondan a artículos publicados en ScienceDirect, estarán directamente disponibles y accesibles desde un mismo punto de acceso.

4.2.8. Análisis DAFO

A partir de los datos recopilados durante la elaboración del informe sobre la UOC, se realiza un análisis DAFO para valorar debilidades, amenazas, fortalezas y oportunidades, de la UOC en relación con la puesta en marcha de un proyecto de *data curation*.

Fortalezas <ul style="list-style-type: none"> CRIS - GIR Existencia de Personal Experimentado en repositorios Repositorio institucional O2 Apoyo de la OSRT Mandato Acceso Abierto UOC Existencia previa en la biblioteca de un plan de formación Predisposición de entrevistas con investigadores Personal acostumbrado a trabajar con datos digitales y conocimiento TIC Grupo de Biblioteca de Apoyo a la Investigación Posicionamiento del Vicerrectorado de Investigación e Innovación Inminente cambio de rectorado? Apoyo del Gabinete Jurídico 	Debilidades <ul style="list-style-type: none"> Bajo nivel de consciencia y conocimiento sobre la gestión de datos por parte del personal científico Ausencia de Personal Experimentado en data curation Equipamiento tecnológico Inexistencia de planes de gestión de datos
Oportunidades <ul style="list-style-type: none"> Posibles cambios normativos a nivel europeo FP8 Existencia de proyectos e iniciativas data curation en Europa Disposición de multitud de herramientas online Existencia de iniciativas de data curation por parte de editoriales Existencia de experiencias próximas Nueva Ley de la Ciencia 	Amenazas <ul style="list-style-type: none"> Inexistencia de políticas de preservación de datos de investigación a nivel estatal Ausencia de un organismo estatal o autonómico que lidere actividades de data curation Ausencia de posicionamiento por parte de las agencias evaluadoras de la investigación Momento de crisis económica: dificultad para encontrar subvenciones

Figura 11. DAFO de los datos de investigación en la UOC

5. Propuestas de actuación

A partir del informe sobre la situación de los datos de investigación en la UOC, se proponen las siguientes medidas:

- I. La creación de un **grupo de trabajo** que, bajo el liderazgo de un grupo del personal de la Biblioteca, elabore un plan director que sienta las bases de una estrategia para la futura implementación de servicios de gestión y preservación de datos de investigación.

El grupo estaría constituido por, al menos, los siguientes agentes: la dirección y un bibliotecario del grupo de Biblioteca de Apoyo a la Investigación, un bibliotecario especialista en el repositorio institucional, un representante del Vicerrectorado de investigación, dos representantes de los grupos de investigación reconocidos por la Generalitat, un técnico del grupo de Aplicaciones y Procesos, y un representante de la OSRT.

- II. El desarrollo de un **Plan Director**, que tenga como objetivo impulsar la cultura de la gestión y preservación de datos de información en la UOC y establecer una guía de las tareas a llevar a cabo por la universidad para en un futuro desarrollar un plan de acción de *data curation*.

El desarrollo de este plan debe alinearse con las líneas estratégicas de la universidad para el período 2010-2014, en el que destaca el eje de R+I+D+T de “Potenciar el desarrollo, la innovación y la transferencia del conocimiento con relación a los ámbitos de investigación” (Plan Estratégico 2010-2014 Universitat Oberta de Catalunya, 2009).



Figura 12 Factores críticos del Plan Director

Para el planteamiento de la elaboración de un plan de acción y de implementación de algún sistema de preservación de datos de investigación se han tenido en cuenta las recomendaciones de organismos como el DCC, guías y experiencias de otras instituciones como la *University of Northampton* (Pickton y otros, 2012), y la *University of Bath* (Ball, 2010) así como el libro de Harvey (2010) y Pryor (2012).

Se han definido cuatro puntos de partida clave:

- 1) En primer lugar, el de los **investigadores**: hay que promover la importancia de la gestión de los datos en el seno del trabajo del investigador. Existe una tarea de divulgación y fomento de esta actividad que debe concluir con la definición de un proceso de recogida de datos unificado para la comunidad científica.
- 2) En segundo lugar, el Plan Director debe apoyarse en la orientación a **nuevos servicios**, gran parte de ellos posiblemente a implementar desde la biblioteca, en el camino hacia la definición de tipologías de datos y el establecimiento de nuevos servicios orientados al usuario investigador. Para ello, será necesaria una nueva perspectiva de la formación y el apoyo

tecnológico que se requiera para definir nuevas herramientas.

- 3) En tercer lugar, resulta clave encajar el Plan Director con el **eje estratégico** definido para las actividades de investigación de la UOC hasta el 2014. Este enclave resulta imprescindible para poder involucrar a todos los equipos necesarios, incluidos los órganos de dirección de la universidad, especialmente a nivel de vicerrectorado y de dirección de programas de doctorado.
- 4) Finalmente, será fundamental atender a la **cultura organizacional** de la UOC, para lograr que, a base de debate y participación entre los distintos agentes involucrados, acepten la corresponsabilidad de llevar a cabo esta tarea.

A continuación se desarrollan algunas propuestas que habrá que debatir y diseñar en las reuniones del grupo de trabajo del Plan Director:

❖ **Línea de acción I: los Investigadores:**

- ✓ Realización de reuniones y entrevistas con la comunidad científica de la UOC. Con el objetivo de comprender las necesidades reales del investigador y trabajar directamente con él, creando sinergias y confianza en la biblioteca para que ellos mismos participen activamente en la creación de un futuro plan de preservación de datos. Para esta acción se propone utilizar una adaptación de la aplicación de pautas de entrevista con investigadores definidas por el recurso *Data Curation Profiles Toolkit*. Esta participación puede resultar especialmente valiosa para, por ejemplo, definir cuáles son los datos más relevantes de su propia investigación, y para definir el mejor modo de llevar a cabo su recolección y preservación.
- ✓ Solicitar un proyecto APLICA de la UOC como fuente de financiación para poder llevar a cabo la creación un plan piloto de recogida de datos de investigación. Los proyectos APLICA están diseñados por la UOC para fomentar el carácter innovador de la universidad, se publican en formato de

convocatorias anuales internas de financiación y están dirigidos a todo el personal de la universidad. La obtención de una ayuda APLICA resuelve, al menos parcialmente, el problema de presupuesto para conseguir desarrollar un proyecto de recogida de datos que sirva para catalogar al máximo nivel de concreción posible las tipologías de datos y formatos creados por los procesos de investigación de la UOC.

- ✓ Desarrollar un plan de difusión y marketing centrado en los investigadores de la UOC que les permita entender la importancia del *data curation* para el futuro de su propia actividad científica. Para esta acción se propone realizar un plan de trabajo transversal con los responsables de las áreas de comunicación de la UOC, la dirección de los distintos estudios, la OSRT y la misma biblioteca. Se trata de trasladar la visibilidad, fomentar y explicar de forma pedagógica a la comunidad científica de la universidad, los beneficios de una correcta gestión y preservación de los datos de investigación. Algunas propuestas irían en la dirección de participar directamente en seminarios, conferencias, y en las Jornadas de investigación que se organizan periódicamente en la universidad. Se plantea invitar a investigadores de otras universidades e instituciones que hayan puesto en práctica acciones de *data curation* a distintos actos de la UOC para que compartan sus experiencias.

❖ Línea de acción 2: la Biblioteca

- ✓ Puesta al día de las directrices y recomendaciones se vayan publicando *sobre la preservación* de datos de investigación tanto a nivel nacional como internacional. Participar, sobre todo desde la biblioteca, de todas aquellas actividades, jornadas, congresos, lecturas, que estén relacionadas con el *data curation*. Para conseguirlo se propone además una reserva del presupuesto de la partida de la biblioteca para poder asistir a estas actividades.
- ✓ Elaborar las bases de un nuevo plan de formación en *data curation*, tanto a nivel de nuevas competencias para el personal bibliotecario como dirigido al investigador. La nueva formación del bibliotecario de datos debe tener en

cuenta la adquisición de competencias que permitan la creación de nuevas políticas y servicios; realizar un análisis de los datos para decidir qué es lo que debe ser conservado; aconsejar sobre la creación y reutilización de los datos; conocer los procesos de almacenaje en repositorios, asegurar la calidad de los datos, así como su correcta estructura y promover la interoperabilidad. El conocimiento de identificadores y metadatos, formatos y la familiaridad con sistemas de *open source* son algunas de las nuevas competencias básicas a conseguir implementar en la biblioteca. Para este punto de acción también se reservará parte del presupuesto de formación anual de la Biblioteca, para asegurar que se dispone de fondos para cursos.

- ✓ Elaborar un proyecto piloto de un nuevo servicio de apoyo al investigador centrado en el consejo y recomendaciones para la gestión y preservación de los datos de investigación. Una vez adquiridas las competencias necesarias mediante el plan de formación mencionado en el punto anterior, la dirección del grupo de biblioteca de apoyo a la investigación debe realizar un estudio sobre la viabilidad, esto es si se dispone de recursos humanos y económicos suficientes para desarrollar un nuevo servicio, y la necesidad por parte del investigador de utilizar dicho servicio.
- ✓ Desarrollar una estrategia interna para promover un plan de *data curation* y un mandato o política de datos de investigación. Desde la dirección de la Biblioteca virtual y de la dirección de grupo SBR se incluirá la promoción del *data curation* en el plan estratégico de la biblioteca para los próximos años y se establecerán reuniones pertinentes para obtener el visto bueno de los organismos de gobierno de la universidad.
- ✓ Realización de un presupuesto, conjuntamente con Aplicaciones y Procesos para poder plantear un proyecto de *data curation*. Antes de plantearse la posibilidad de llevar a cabo un proyecto propio de almacenamiento de datos de investigación en la UOC, será necesario realiza un estudio técnico, entre la biblioteca y los técnicos de Aplicaciones y Procesos, para evaluar la adecuación o no de una herramienta de este tipo a la universidad.

❖ Línea de acción 3: la Universidad

- ✓ Potenciar la elaboración de una normativa o mandato institucional relativo a la gestión y preservación de los datos de investigación a través del Vicerrectorado de Innovación e Investigación. Para esta acción se seguirán las directrices propuestas por el Data Curation Center, y se solicitará la participación de la biblioteca y de algún agente externo a la universidad, por ejemplo, de documentalistas de otras instituciones que ya tengan experiencia en la redacción de este tipo de documentos.
- ✓ Participar activamente en la creación de algún grupo de trabajo a nivel estatal o autonómica para afrontar el tema de data curation conjuntamente con otras instituciones académicas que se estén planteando el mismo problema (por ejemplo participar en el grupo de Repositorios de la Fecyt).
- ✓ Entablar conversaciones y reuniones con el equipo de tecnología de biblioteca para valorar conjuntamente la posibilidad de aplicar o implementar alguna herramienta TIC que facilite la gestión y preservación de los datos de investigación creados por la UOC, o estudiar las posibilidades que pueda ofrecer el O2. Esta acción se desarrolla en paralelo con la del último punto de la línea estratégica de biblioteca.
- ✓ Establecer lazos estables para conseguir la inclusión del desarrollo de un plan de acción de *data curation* en los planes del consejo de gobierno de la universidad. El contexto de crisis económica actualmente no facilita esta tarea, pero desde la biblioteca y desde la comunidad científica se debe hacer todo lo posible para que el tema de los datos de investigación se incluya en las agendas de la dirección de la universidad. Para este cometido será necesario el compromiso de la dirección de todos los agentes implicados.

6. Conclusiones

El estudio del contexto general y del marco teórico del *data curation* a nivel internacional muestra que a pesar de ser campo de estudio novedoso y en proceso de cambios y actualización permanente, cuenta con muchas organizaciones y centros que se encargan de estudiar y desarrollar la materia.

Los resultados de la realización de los primeros apartados de este trabajo ofrecen las siguientes conclusiones:

1. Se confirma la existencia de una gran comunidad de *data curators* a nivel internacional liderado desde los países anglosajones que se contrapone a una realidad ausente de organismos, iniciativas o apoyos del entorno español. Ésta situación ofrece a las bibliotecas universitarias, centros de investigación y organismos de evaluación nacionales la oportunidad de tomar las riendas necesarias para emprender el liderazgo del *data curation* a nivel español.

2. Se abre un nuevo nicho de mercado en materia de e-ciencia desde el punto de vista de las bibliotecas académicas. Los esfuerzos de adaptación al nuevo contexto de la investigación, donde los datos reciben más protagonismo, representan un auténtico reto para estos grupos de apoyo a la investigación.

Por otro lado, la realización de un estudio práctico sobre la situación de los datos de investigación sobre el caso de la Universitat Oberta de Catalunya, permite obtener una fotografía completa de su posición y se concluye lo siguiente:

3. El actual contexto de crisis económica, sumado a la situación de la propia universidad en la que hasta el momento se ha detectado una falta de interés hacia el tema por parte de los órganos de gobierno, dificultan la posibilidad de iniciar un plan de *data curation*.

4. El informe y las acciones propuestas plantean la existencia de opciones alternativas para la mejora de la gestión y de la preservación de datos de investigación. Sin tener que llegar a la necesidad de un campus de investigación que facilite el desarrollo de la e-ciencia, existe una fase previa relativa al

almacenamiento, conservación y preservación de los datos de investigación que sí que podría ser ejecutada. Con el esfuerzo de la Biblioteca de la UOC y mediante la colaboración con otros grupos de la universidad cercanos a los investigadores, se hace posible la implementación de estas mejoras.

5. Para la Universitat Oberta de Catalunya su experiencia en el campo de los repositorios institucionales y sistemas CRIS la sitúa en una situación más ventajosa. A pesar de que resulta complicado plantearse la posibilidad de conseguir una mejora tecnológica que permita la inclusión de datos de investigación, existe la voluntad y el interés por parte de la biblioteca en detectar las nuevas necesidades de los investigadores, siendo éste un factor clave para llevar a cabo la implementación de servicios de *data curation*; así lo demuestra la existencia de estudios recientes sobre el papel del bibliotecario de investigación (Riera Quintero, Padrós Cuxart y Zuñiga Ruiz, 2012).

6. La creación de políticas o mandatos institucionales relacionados con dichos datos seguida por una inversión en la formación de bibliotecarios e investigadores resulta esencial para plantearse cualquier propuesta de mejora de la preservación de los datos de investigación. La ausencia de un procedimiento mínimo común para todos los datos generados y utilizados en el transcurso de la investigación de los distintos grupos, aparte de las implicaciones relativas a la seguridad de los datos, crea un vacío que no alienta a su más adecuada conservación. Es necesario establecer unos criterios mínimos que faciliten la conservación de los datos primarios generados en el transcurso de la investigación, y en cuya definición deben participar todas las partes implicadas de la UOC.

En definitiva, el acercamiento al *data curation* muestra como los bibliotecarios deben ponerse, una vez más, en la piel del mejor de los estrategias para llevar a cabo una “evangelización” de *data curation* a vicerrectorados, investigadores y órganos de decisión de la universidad sobre la importancia de impulsar cierta obligatoriedad o normativa para que las bibliotecas puedan seguir ejerciendo su papel de mediadora con los distintos procesos de investigación. Con la única finalidad de asegurar la preservación de los datos de investigación de una

institución, potenciar y apoyar iniciativas que, a nivel estatal o internacional, reconozcan esta necesidad y ayuden a la creación de directrices comunes que permitan a las universidades, pese a la falta de recursos económicos y/o tecnológicos, implementar acciones de mejora para la preservación de los datos de investigación.

Bibliografía

- Adoració Pérez, Jordi Serrano & Marta Enrech. (1997). La Biblioteca Virtual de la Universitat Oberta de Catalunya: un nuevo concepto de biblioteca. Barcelona. Retrieved from <http://openaccess.uoc.edu/webapps/o2/bitstream/10609/332/1/1727.pdf>
- Altenhöner, R. (2012). Libraries as service-brokers for digital data curation: Practical insights from the DFG project DP4lib (Digital Preservation for libraries). *78th IFLA World Library and Information Congress*. Helsinki. Retrieved from <http://conference.ifla.org/sites/default/files/files/papers/wlic2012/116-altenhoner-en.pdf>
- Angevaere, I. (2009). Curating Research: e-Merging New Roles and Responsibilities in the European Landscape (Conference Report). *LIBER Quarterly* (Vol. 19, pp. 180–191). Retrieved from <http://liber.library.uu.nl/index.php/lq/article/view/URN:NBN:NL:UI:10-1-113564/8250>
- Ball, A. (2010). Review of the State of the Art of the Digital Curation of Research Data. (E. P. Document, Ed.). Bath, UK. Retrieved from <http://opus.bath.ac.uk/19022/2/erim1rep091103ab12.pdf>
- Barateiro, J., Antunes, G., Cabral, M., Borbinha, J. & Rodrigues, R. (2008). Digital Preservation of Scientific Data. (B. Christensen-Dalsgaard, D. Castelli, B. Ammitzbøll Jurik, & J. Lippincott, Eds.) *Research and Advanced Technology for Digital Libraries*. Springer Berlin / Heidelberg. Retrieved from http://0-dx.doi.org.cataleg.uoc.edu/10.1007/978-3-540-87599-4_41
- Beagrie, N., Chruszcz, J., & Lavoie, B. (2008). *Keeping Research Data Safe: A Cost Model and Guidance for UK Universities. Education* (p. 169). London: JISC.

Retrieved from

<http://www.jisc.ac.uk/media/documents/publications/keepingresearchdatasafe0408.pdf>

Board, N. S. (2005). *Long-Lived Data Collections: Enabling Research and Education in the 21st Century*. Arlington: National Science Foundation. Retrieved from <http://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf>

Borgman, C. L. (2007). *Scholarship in the digital age: information, infrastructure, and the Internet*. Cambridge, Mass: MIT Press.

Charles W. Bailey, J. (2012). Digital Curation Bibliography: Preservation and Stewardship of Scholarly Works. Retrieved from <http://digital-scholarship.org/dcpb/dcb.htm>

Christensen-Dalsgaard, B. (2012). *Ten recommendations for libraries to get started with research data management* (p. 3). Retrieved from [http://www.libereurope.eu/sites/default/files/The research data group 2012 v7 final.pdf](http://www.libereurope.eu/sites/default/files/The%20research%20data%20group%202012%20v7%20final.pdf)

Corall, S. (2008). Research Data Management: Professional Education and Training Perspectives. Manchester: DCC. Research Data Management Forum. Retrieved from <http://www.dcc.ac.uk/sites/default/files/documents/RDMF/RDMF2/07Corrall.pdf>

Council, N. R. (2009). *Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age*. The National Academies Press. Retrieved from http://www.nap.edu/catalog.php?record_id=12615

Daniel Torres-Salinas. (2010). Compartir datos (data sharing) en ciencia: el contexto de una oportunidad. *ThinkEPI*. Retrieved from

<http://www.thinkepi.net/compartir-datos-data-sharing-en-ciencia-el-contexto-de-una-oportunidad>

Donnelly, M. (2008). RDMF2: Core Skills Diagram. *DCC.Research Data Management Forum*. Retrieved August 20, 2012, from <http://data-forum.blogspot.com.es/2008/12/rdmf2-core-skills-diagram.html>

Edinburgh, T. U. of. (2011). Research data management guidance.

España. (2011). Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación. Boletín Oficial del Estado.

Evans, J., Cole, G., & Lloyd-Jones, H. (2012). Postgraduate Research Data : a new type of challenge for repositories? Edinburgh: University of Exeter, United Kingdom.

Generalitat de Catalunya. (1995). Llei 3/1995, de 6 d'abril, de reconeixement de la Universitat Oberta de Catalunya. Retrieved from http://www.gencat.cat/diue/doc/doc_59501855_1.pdf

GrandIR. (2012). Jornada técnica "Avances en la gestión de datos de investigación en España": Programa. Retrieved from <http://www.grandir.com/es/sesiones-tecnicas/avances-en-la-gestion-de-datos-de-investigacion-en-espana/programa>

Gutmann, M. P., Abrahamson, M., Adams, M. O., Altman, M., Arms, C. R., Bollen, K., Carlson, M., et al. (2009). From Preserving the Past to Preserving the Future: The Data-PASS Project and the challenges of preserving digital social science data. *Library Trends*, 57, 315–337.

Harvey, R. (2010). *Digital Curation: a How To Do It Manual*. Neal-Schuman Publishers.

- Heidorn, P. B. (2011). The Emerging Role of Libraries in Data Curation and E-science. *Journal of Library Administration*, 51(7-8), 662– 672.
doi:10.1080/01930826.2011.601269
- High level Expert Group on Scientific Data. (2010). *Riding the wave: How Europe can gain from the rising tide of scientific data*. Retrieved from http://ec.europa.eu/information_society/newsroom/cf/document.cfm?action=display&doc_id=707
- Joint, N. (2007). Data preservation, the new science and the practitioner librarian. *Library Review*, 56(6), 451–455.
- Jones, S. (2009). *A report on the range of policies required for and related to digital curation*. Glasgow,. Retrieved from http://www.dcc.ac.uk/sites/default/files/documents/reports/DCC_Curation_Policies_Report.pdf
- Jones, S., Ball, A., & Ekmekcioglu, C. (2008). The Data Audit Framework: A First Step in the Data Management Challenge. *International Journal of Digital Curation*, 3(2), 112–120. Retrieved from <http://eprints.erpanet.org/159/>
- Jones, S., Ross, S., & Ruusalepp, R. (2008). The Data Audit Framework: a toolkit to identify research assets and improve data management in research led institutions. *British Library Conference Centre 2930 Sept 2008*, 225–231. Retrieved from <http://eprints.gla.ac.uk/6240/>
- Keefer, A. (2011). La preservación de los datos de investigación y las agencias de financiación de la I+D. *Blok de BiD*. Retrieved from <http://www.ub.edu/blokdebid/es/node/130>
- Keralis, S. D. C. K. (2012). Data Curation Education: A Snapshot — Council on Library and Information Resources. *The Problem of Data*. Retrieved from <http://www.clir.org/pubs/reports/pub154/education>

- Lewis, M. J. (2010). Libraries and the management of research data. In Sue McKnight (Ed.), *Envisioning Future Academic Library Services* (pp. 1–28). Facet Publishing. Retrieved from <http://www.facetpublishing.co.uk/title.php?id=691-6>
- López-Pérez, C., & Vaquer-Suñer, C. (2011). Accés obert i el repositori institucional de la UOC (O2). Universitat Oberta de Catalunya. Retrieved from <http://openaccess.uoc.edu/webapps/o2/handle/10609/8369>
- Martinez-Urbe, L., & Macdonald, S. (2008). Un nuevo cometido para los bibliotecarios academicos: data curation. *El Profesional de la información*, 17(3), 273.
- Ogburn, J. L. (2010). The Imperative for Data Curation, 10(2), 241–246. Retrieved from http://muse.jhu.edu/login?auth=0&type=summary&url=/journals/portal_libraries_and_the_academy/v010/10.2.ogburn.html
- Osswald, Achim & Strathmann, S. (2012). The Role of Libraries in Curation and Preservation of Research Data in Germany: Findings of a survey. *78th IFLA World Library and Information Congress*. Helsinki. Retrieved from <http://conference.ifla.org/sites/default/files/files/papers/wlic2012/116-osswald-en.pdf>
- Pickton, Miggie; Jones, Sarah & Guy, M. (2012). Leading from the library: data management initiatives at the University of Northampton. *78th IFLA World Library and Information Congress*. Helsinki. Retrieved from <http://conference.ifla.org/sites/default/files/files/papers/wlic2012/116-pickton-en.pdf>

- Prieto, T. (2012). Experiencias de interoperabilidad entre CRIS y repositorios en Catalunya. Universitat Politècnica de Catalunya. Retrieved from <http://openaccess.uoc.edu/webapps/o2/handle/10609/10881>
- Pryor, G. (2012). *Managing research data* (p. 239). London: Facet Publishing.
- Ramírez, M. L. (2011). Opinion: Whose role is it anyway?: A library practitioner's appraisal of the digital data deluge. *Bulletin of the American Society for Information Science and Technology*, 37(5), 21–23. Retrieved from http://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1080&context=lib_fac
- Reilly, S. (2012). The role of libraries in supporting data exchange. *78th IFLA World Library and Information Congress*. Helsinki. Retrieved from <http://conference.ifla.org/sites/default/files/files/papers/wlic2012/116-reilly-en.pdf>
- Riera Quintero, C., Padrós Cuxart, R., & Zuñiga Ruiz, A. (2012). Bibliotecarios de investigación en un entorno virtual: nuevas habilidades, nuevos servicios. Barcelona. Retrieved from <http://openaccess.uoc.edu/webapps/o2/handle/10609/16141>
- Rodríguez Miranda, Alvaro ; García Gómez, Consol & de Castro, P. (2012). Gestión de datos de investigación en repositorios de acceso abierto: una visión panorámica y un caso práctico en la UPV/EHU. Bilbao: OSREP5.
- Sallans, A. L., Sherry; (2011). How to Re-Tool Librarians for Data Curation. a, Charlottesville, Virginia, U.S.A: Scientific Data Consulting Group, Charles L. Brown Science and Engineering Library. Retrieved from http://www.dcc.ac.uk/webfm_send/297
- Silvia Arano, G. M., Marina Losada, Marta Villegas, Anna, & Casaldàliga y Núria Bel. (2011). La comunidad «Recursos y datos primarios» de la Universitat Pompeu

Fabra: los repositorios institucionales como infraestructuras científicas: estudio de caso. *Revista Española de Documentación Científica*, 34(3), 385–407.

Strasser, C., Cook, R., Michener, W., & Budden, A. (2012). *Primer on Data Management : What you always wanted to know* (pp. 1–11). Retrieved from http://www.dataone.org/sites/all/documents/DataONE_BP_Primer_020212.pdf

Stuart Macdonalda & Luis Martinez-Urbe. (2010). Collaboration to Data Curation: Harnessing Institutional Expertise. - *New Review of Academic Librarianship*, 16(1), 4–16. doi:10.1080/13614533.2010.505823

Swan, A., & Brown, S. (2008). *Skills, Role & Career Structure of Data Scientists & Curators: Assessment of Current Practice & Future Needs. Current Practice* (p. 32 p. ST – Skills, role and career structure of d). Key Perspectives. Retrieved from <http://www.jisc.ac.uk/media/documents/programmes/digitalrepositories/dataskillsreport.pdf>

Tenopir, C., Sandusky, R.J., Allard, S. & Birch, B. (2012). Academic librarians and data research services: preparation and attitudes. *78th IFLA World Library and Information Congress*. Helsinki. Retrieved from <http://conference.ifla.org/sites/default/files/files/papers/wlic2012/116-tenopir-en.pdf>

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M., et al. (2011). Data Sharing by Scientists: Practices and Perceptions. - *PLoS ONE*. Public Library of Science. doi:10.1371/journal.pone.0021101

The Australian National University. (2010). *ANU Data Management Manual*. Retrieved from http://regnet.anu.edu.au/sites/default/files/files/ANU_Data_Management_Manual.pdf

- Torres-Salinas, D. (2010). Primeros pasos hacia la gestión de datos de investigación en las universidades: la iniciativa DAF. *Notas ThinkEPI*. Retrieved from <http://www.thinkepi.net/primeros-pasos-hacia-la-gestion-de-datos-de-investigacion-en-las-universidades-la-iniciativa-daf>
- Treloar, A., Groenewegen, D., & Harboe-Ree, C. (2007). The data curation continuum: Managing data objects in institutional repositories. *D-Lib Magazine*, 13(9), 4.
- UOC. (2009). Plan Estratégico 2010-2014 Universitat Oberta de Catalunya. Retrieved from http://cv.uoc.edu/UOC/a/intrauoc/docs/actualitat/rrhh/docs/PLA_CAST.pdf
- UOC. (2010). *Política institucional de acceso abierto de la Universitat Oberta de Catalunya*. Universitat Oberta de Catalunya. Retrieved from <http://openaccess.uoc.edu/webapps/o2/handle/10609/4967>
- Uniersitat Oberta de Catalunya. (2010). Servicios de la Biblioteca para la investigación. Retrieved from <http://hdl.handle.net/10609/8982>
- UPV/EHU. (2012). ADDI: Repositorio Institucional de la Universidad del País Vasco. Retrieved from <https://addi.ehu.es/>
- Uhlir, P. F., & Schröder, P. (2007). Open Data for Global Science. *Data Science Journal*, 6, OD36–OD53.
- VV.AA. (2012). *Jornada GrandIR sobre iniciativas para la gestión de datos de investigación en España UPC-Campus Nord BCN, 10 de mayo 2012*. Retrieved from <http://www.grandir.com/es/descargas?download=55:sesion-rdm-barcelona-resumen-y-conclusiones&start=20>
- Walters, T. O. (2009). Data Curation Program Development in U.S. Universities: The Georgia Institute of Technology Example. *The International Journal of Digital*

Curation, 4(3), 83–92. Retrieved from
<http://ijdc.net/index.php/ijdc/article/view/136/0>

Wilson, J. A. J., Martinez-Urbe, L., Fraser, M. A., & Jeffreys, P. (2011). An Institutional Approach to Developing Research Data Management Infrastructure. *International Journal of Digital Curation*, 6(2), 274–287.

Zimmermann, R. (2012). Preparatory work for the Framework Programme 8 in Software and Services research. Retrieved from
<http://cordis.europa.eu/fp7/ict/ssai/docs/fp8-preparations/rainerzimmermann.pdf>

Anexo: Listado de instituciones, organismos y páginas web consultadas

- 3TU.Datacentrum [<http://datacentrum.3tu.nl/en/home/>]
- Arts and humanities Data Service (AHDS) [<http://www.ahds.ac.uk/>]
- Arts and Humanities Research Council (AHRC)
[<http://www.ahrc.ac.uk/Pages/Home.aspx>]
- Archaeology Data Service (ADS) [<http://archaeologydataservice.ac.uk/>]
- Australian National Data Service (ANDS) [<http://ands.org.au/>]
- Biotechnology and Biological Sciences Research Council (BBSRC)
[<http://www.bbsrc.ac.uk/home/home.aspx>]
- Cornell University [<http://www.cornell.edu/>]
- Data Liberation Initiative [<http://www.statcan.gc.ca/start-debut-eng.html>]
- Data Curation Profiles Toolkit [<http://datacurationprofiles.org/>]
- Digital Curation Centre (DDC)[<http://www.dcc.ac.uk/>]
- Digital Curation Center Resources [<http://www.dcc.ac.uk/resources>]
- DISC-UK DataShare [<http://www.disc-uk.org/datashare.html>]
- Distributed Data Curation Center (D2C2) [<http://d2c2.lib.purdue.edu/>]
- Digital Curation Exchange [<http://digitalcurationexchange.org/>]
- Digital Curator Vocational Education Europe(DigVurV) [<http://www.digcur-education.org/eng>]
- Dryad [<http://datadryad.org/>]
- European Data Infraestructure (EUDAT) [<http://www.eudat.eu/>]
- GrandIR [<http://www.grandir.com/>]
- Inter-university Consortium for Political and Social Research (ICPSR)
[<http://www.icpsr.umich.edu/icpsrweb/landing.jsp>]
- Joint Information Systems Committee (JISC)[<http://www.jisc.ac.uk/>]
- London School of Economics [<http://www2.lse.ac.uk/home.aspx>]
- National Academy of Science (NAS) [<http://www.nasonline.org/>]
- National Science Board (NSB) [<http://www.nsf.gov/nsb/>]
- National Science Foundation (NSF) [<http://www.nsf.gov/>]

- Network of expertise in Long-Term Storage (NESTOR)
[<http://www.digitalpreservationeurope.eu/competence-centres/list/?id=92>]
- Massachussets Institute of Technology (MIT)[<http://www.mit.edu/>]
- MIT Libraries: Data Management and Publishing
[<http://libraries.mit.edu/guides/subjects/data-management/plans.html>]
- Purdue University Library [<http://www.lib.purdue.edu/>]
- UK Data Archive [<http://www.data-archive.ac.uk/>]
- University of Cambridge [<http://www.cam.ac.uk/>]
- University of Edinburgh [<http://www.ed.ac.uk/home>]
- University of Illynois [<http://illinois.edu/>]
- University of New Mexico [<http://www.unm.edu/>]
- Universitat Oberta de Catalunya [<http://www.uoc.edu/portal/es/index.html>]
- Universitat Politècnica de Catalunya [<http://www.upc.edu/>]